

**Models for Two-state Disease Processes with
Applications to Relapsing-Remitting Multiple
Sclerosis**

by

Jochen Brumm

Hauptdiplom (Mathematics), Universitaet Bonn, 1997

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE STUDIES

(Department of Statistics)

we accept this thesis as conforming
to the required standard

The University of British Columbia

January 2000

© Jochen Brumm, 2000

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of STATISTICS

The University of British Columbia
Vancouver, Canada

Date January 27, 2000

Abstract

In diseases like relapsing-remitting multiple sclerosis (MS), patients experience repeated transitions between symptom-free and symptomatic disease states (the symptomatic state is called an exacerbation). Analyses for this kind of data commonly ignore the information available on the second state (the lengths of the exacerbations, for example).

In this thesis, we consider models that incorporate the second state into the analyses. The basic stochastic models are Markov chains, alternating renewal processes and marked point processes. For the Markov chains and alternating renewal process models, we consider simple fixed effects models as well as random effects models where the random effects are introduced to allow for heterogeneity between patients and correlation of data on one patient. For these models, the statistical inference is based on maximum likelihood. For the marked point process model, we use a generalized estimating equation approach.

We apply these models to a data set from a MS clinical trial. The aim of the analyses is to relate the available covariates to the disease process. We do not attempt a comprehensive analysis of the data set, rather the aim here is to see what can be achieved and which questions can be addressed with the different models.

Contents

Abstract	ii
Contents	iii
List of Tables	vi
List of Figures	viii
Acknowledgements	x
1 Introduction	1
1.1 Motivation for this Thesis	1
1.2 Description of the Data	3
1.2.1 Design of the Clinical Trial	3
1.2.2 Baseline Covariates	3
1.2.3 Outcome Variables	6
1.3 Previous Analyses and Our Objectives	11
1.4 Models Considered	13
1.4.1 Markov Chains	13
1.4.2 Alternating Renewal Processes	14
1.4.3 Marked Point Processes	15
1.5 Methods of Analysis	16
1.5.1 Markov Chains	17
1.5.2 Alternating Renewal Processes	18
1.5.3 Marked Point Processes	18
1.6 Outline of this Thesis	19

2	A Markov-Chain Model	20
2.1	Introduction	20
2.2	The Likelihood	21
2.3	A Parametric Fixed Effects Model	23
2.4	A Random Effects Model	24
2.5	Model Assessment	26
2.6	Application to the Exacerbation Data	28
2.6.1	Fixed Effects Models	29
2.6.2	Random Effects Models	35
2.7	Discussion	36
3	Point Processes: Basic Properties and Preliminary Analyses	41
3.1	Terminology and Basic Properties	42
3.2	Initial Data Analysis	45
4	An Alternating Renewal Process Model	52
4.1	Introduction	52
4.2	Fixed Effects	54
4.2.1	Parameterization of the Intensities	54
4.2.2	The Likelihood	54
4.3	Random Effects	55
4.3.1	The Distribution of the Random Effects	56
4.3.2	The Likelihood	57
4.4	Parameterization of the Baseline Intensity	58
4.5	Numerical Methods	61
4.5.1	Random Effects	61
4.5.2	Splines	62
4.6	Comments on the Numerical Method	63
4.7	Model Assessment	65
4.8	Results for the Alternating Renewal Approach	67
4.8.1	Models Fit	67
4.8.2	Fixed Effects	68
4.8.3	Random Effects	72
4.9	Discussion	76

5	A Marked Point Process Model	82
5.1	Introduction	82
5.2	Estimating Equations for Poisson Processes	84
5.3	Generalized Estimating Equations for Point Processes	85
5.4	Estimating Equations for Marked Poisson Processes	88
5.5	GEE for Marked Point Processes	91
5.6	Parametrization of the Mean Functions	92
5.7	Application to the MS Dataset	95
5.8	Discussion	96
6	Discussion	99
6.1	Conclusions for the Data Set	99
6.2	Comparison of Methods	103
6.3	Conclusion	105
	Bibliography	106

List of Tables

1.1	Number of male and female patients in each treatment group .	5
2.1	Covariate effects for the first-order fixed effects model (2.5) .	31
2.2	Covariate effects for the first-order fixed effects model, separate regressions for the transitions	31
2.3	Z-scores for covariate effects for the second-order fixed effects model (see (2.6)), initial model	33
2.4	Covariate effects for the reduced (after backward selection) second-order fixed effects model (2.6), final model; a — indicates this term was dropped from the model	34
2.5	Estimates for the second-order transition probabilities from the reduced second-order model (other estimates omitted)	34
2.6	Covariate effects for the random intercept models (2.8) and (2.9). σ denotes the standard deviation for the random intercepts	35
2.7	Comparison of covariate effects for first-order models	37
2.8	Comparison of z-scores for first-order models	37
3.1	Number of exacerbations by treatment and time-period	49
4.1	Quartiles for the time-scale parametrizations (in days)	68
4.2	Results for the fixed effects models	69
4.3	Results for the random intercepts models	75
4.4	Estimated effects of baseline covariates for fixed effects and random intercept models for the model with a piecewise constant parametrization of the time-scales	78
4.5	Z-scores for different models	79
5.1	Estimated effects for the mean counts of exacerbations	95

5.2	Estimated standard errors for the mean counts of exacerbations	95
5.3	Estimated effects for the mean exacerbation lengths	97
5.4	Estimated standard errors for the mean exacerbation lengths .	97
6.1	Z-scores for different simple models (fixed-effects and naive standard errors)	101
6.2	Z-scores for different models (random effects and robust standard errors)	102

List of Figures

1.1	Kaplan-Meier survival curves for time on study	4
1.2	Boxplots of age, time since onset and EDSS by treatment group	5
1.3	Patterns of exacerbations for selected patients	7
1.4	Length of exacerbation-free periods by treatment group	8
1.5	Length of exacerbations by treatment group	8
1.6	Proportion of time in exacerbation over time in 200 day periods by treatment	10
1.7	Proportion of time in exacerbation versus covariates	10
2.1	No exacerbation to exacerbation transition probabilities by time period	30
2.2	Exacerbation to exacerbation transition probabilities by time period	30
2.3	Average estimated probability and proportion of patients in ex- acerbation over time, by quartiles of estimated transition prob- ability (see text for details)	39
2.4	Average estimated probability and proportion of patients in ex- acerbation over time	40
3.1	Rate of exacerbations versus covariates	46
3.2	Average lengths of exacerbations versus covariates	47
3.3	Rate of exacerbation over time in 200 day time periods by treat- ment	50
3.4	Mean and median average lengths of exacerbations over time in 200 day time periods by treatment	51
4.1	Integrated intensity versus number of exacerbations	65

4.2	g for a typical and an extreme case (see text for definition) for different values of ω	66
4.3	Boxplots of Markov and renewal times (in units of 100 days) .	70
4.4	Estimated effect of time-scales for transition into exacerbation with pointwise confidence intervals, fixed effects model	71
4.5	Estimated effect of time-scales for transition out of exacerbation with pointwise confidence intervals, fixed effects model	72
4.6	Variance versus mean for the lengths of inter-event times (in 100 days) for individual patients	73
4.7	Mean sojourn time (in 100 days) in state 1 versus state 2 . . .	74
4.8	Profile log-likelihood for independent random effects	74
4.9	Estimated time-scales for transition into exacerbation with pointwise confidence interval, random intercept model	77
4.10	Estimated time-scales for transition out of exacerbation with pointwise confidence interval, random intercept model	77

Acknowledgements

I would like to thank my supervisor, John Petkau, for his guidance on this project and his encouragement to participate in the opportunities that presented themselves. I am grateful to Paul Gustafson for his careful reading of the manuscript. I also want to thank Michael Brauer for his generous support and Ruben Zamar for encouraging me to pursue this program. Additionally, I would like to thank the UBC Statistics Department for making my stay enjoyable and my good friend Hari Mailvaganam for his patience.

JOCHEN BRUMM

*The University of British Columbia
January 2000*

Chapter 1

Introduction

1.1 Motivation for this Thesis

The motivation for this thesis comes from a data set collected in a multiple sclerosis (MS) clinical trial. In this section we give some background on MS and this clinical trial and in the next section we describe the data. For a more detailed description of the data set and more information on MS and this trial, see D'yachkova [6].

Multiple sclerosis is a common neurological disease in young adults. It is most prevalent in north-west Europe, in Canada, the northern United States, southern Australia and New Zealand; women are more frequently affected by it than men with a ratio of about 3 to 2. Its cause is unknown; it is characterized by the appearance of lesions in the central nervous system. MS patients experience repeated exacerbations during the course of the disease. The initial exacerbation, which leads to the diagnosis of MS, is usually experienced as a young adult (most commonly in the early 30's). During these exacerbations patients can suffer from a variety of symptoms such as blurred vision, a sensation of numbness in parts of the body or loss of control of the movements in parts of the body.

After the initial exacerbation, MS patients typically enter into the relapsing-remitting stage of the disease, which usually lasts for about 5 years. In this stage, patients experience repeated exacerbations that can vary in length and severity. Initially, patients recover completely or almost completely

from these exacerbations, but the recovery can become less and less complete after each exacerbation. Eventually patients leave the relapsing-remitting stage and move on to the secondary progressive stage, where the degree of disability is relatively stable and the patients' health declines steadily.

The MS disease process is extremely variable. Some patients experience a benign form with complete recovery from the exacerbations; others develop grave disabilities. The exacerbations can also be quite different in length and severity.

MS has many symptoms, and there are a number of ways to evaluate the progression of the disease and hence potentially judge the efficacy of a treatment. These include

- the occurrence of exacerbations
- the degree of disability
- counts of lesions in the brain

The degree of disability for MS patients is commonly measured by the Kurtzke expanded disability score (EDSS). The EDSS is an ordinal score based on a neurological examination that describes the degree of disability of a patient (0 meaning healthy, 10 meaning death from MS). Lesion counts can be obtained using magnetic resonance imaging (MRI); for more detail on MRI and lesion counts, see Smith [29].

To date there is no known cure that can completely reverse all the effects of MS. Interferon-1b is the first drug that was demonstrated to be beneficial to MS patients. The data set to be analyzed in this thesis comes from the clinical trial involving patients with relapsing-remitting MS in which it was first demonstrated that Interferon-1b decreases the rate of exacerbations and increases the proportion of patients remaining exacerbation-free [9]. There have been subsequent trials demonstrating that other drugs (Interferon-1a and Copolymer) are beneficial for relapsing-remitting MS patients, see [11], [27] and [12]. More recently Interferon beta-1b has also been demonstrated to be beneficial for secondary progressive MS patients [8].

1.2 Description of the Data

In this section we describe the design of the trial and the data set to be analyzed in this thesis.

1.2.1 Design of the Clinical Trial

As mentioned above, the data considered in this thesis comes from a clinical trial [9]. This trial involved 372 patients in 11 centers in the U.S. and Canada on three treatment arms (placebo, low dose and high dose). Patients were randomized to these arms within each center. The dosage was 1.6 million international units (MIU) for the low dose treatment and 8 MIU for the high dose treatment. The inclusion criteria for the trial were that patients had to have experienced at least 2 exacerbations in the previous 2 years and their EDSS at time of randomization had to be 5.5 or less.

The original planned duration of this trial was 2 years; it was later decided to extend the trial to 3 years. After 3 years the treatment effect became obvious, so all patients in the study were offered the high dose treatment for 2 more years (see [10] for results on the complete 5 year period). We analyze the data from the 3-year period. Figure 1.1 shows the proportion of patients remaining on study by treatment group, the solid line in the plot indicates the end of the 2-year period. The plot shows that the high dose group had the most early drop-outs, but all groups have a modest number of patients leaving the trial during the first 2 years, with the number of patients remaining in each group approaching 80% after 2 years. There are a number of patients in each treatment arm that drop out at the end of the 2-year period, but the proportion remaining for most of the third year of the study is still about 70% in each group.

1.2.2 Baseline Covariates

The baseline covariates we consider are:

- gender
- age

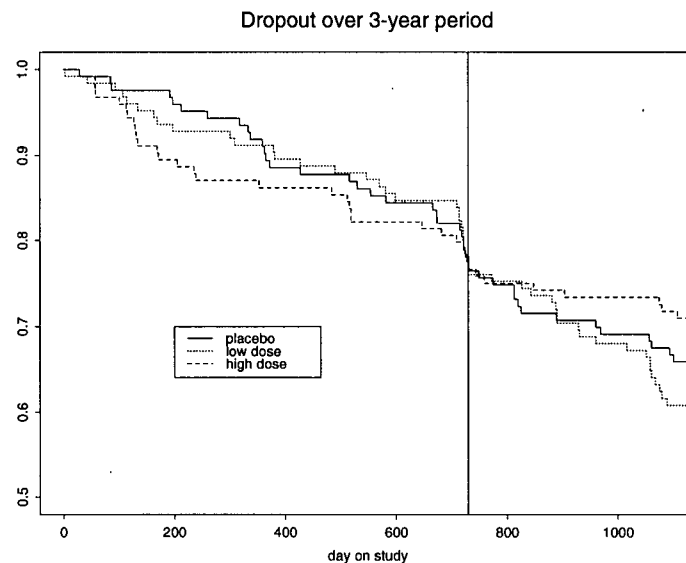


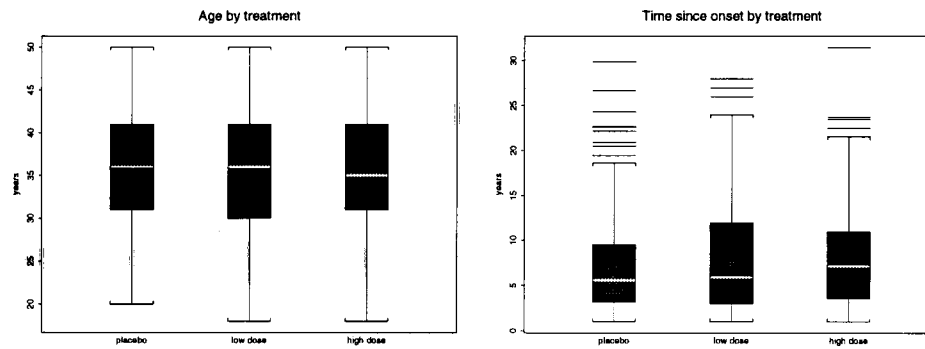
Figure 1.1: Kaplan-Meier survival curves for time on study

- the time since onset of MS
- the Kurtzke Expanded Disability Status Score (EDSS)

The center identities are also available, but we don't incorporate them into the analyses for reasons of convenience (see Section 1.3 for a discussion of the objectives of our analyses). Table 1.1 shows that there are more female than male patients in the trial; the female to male ratio is slightly more than 2 in each treatment arm. The boxplots of age by treatment group in Figure 1.2 show that about half the patients were between 30 and 40 years old, with all patients being between 18 and 50 years old. The distribution of the ages is almost the same for the 3 treatment groups. Figure 1.2 also shows that the time since onset of MS varies between 1 and 31 years, and the median time since onset is slightly higher in the high dose group. The boxplots for the EDSS scores by treatment group (Figure 1.2) shows that the distribution is fairly balanced across the treatment arms, with the scores ranging from 0 to 5.5 in each arm.

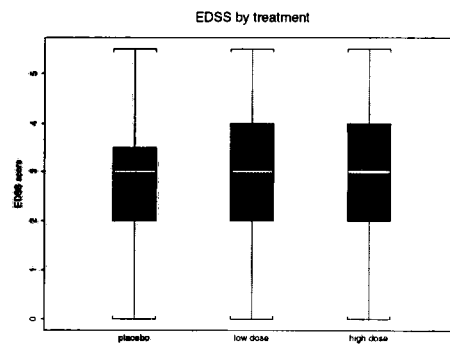
	placebo	low dose	high dose
male	35	40	38
female	88	85	86

Table 1.1: Number of male and female patients in each treatment group



(a) Age by treatment group

(b) Time since onset by treatment group



(c) EDSS by treatment group

Figure 1.2: Boxplots of age, time since onset and EDSS by treatment group

1.2.3 Outcome Variables

The only outcomes we consider are related to the exacerbation data. The start and end dates of each exacerbation were recorded as well as the severity of the exacerbation; our analyses use the data on the start and end of the exacerbations. We don't use the data on severity, although they could contain valuable information; we comment on this in Chapter 6. We only use data from the first three years of the trial; we will discuss how the data enter into the analyses in the respective chapters.

Figure 1.3 shows the patterns of exacerbations for some patients selected for illustrative purposes from the high dose and placebo groups. We chose these patients to demonstrate that the data is quite variable, even within a patient, in both the high dose and placebo groups. This plot shows that the patterns over time are different for different patients. Some patients have exacerbations in close succession, some have no exacerbation at all and others have a succession of early exacerbations followed by a long period without an exacerbation and then a late exacerbation. Another patient has one early exacerbation followed by three late ones after a long exacerbation-free period. The length of time in exacerbation is also quite variable, with some exacerbations lasting only a few days and others lasting almost 100 days.

For an initial summary of the information available on the exacerbations, we use boxplots for the length of times not in exacerbation and of times in exacerbation. Figure 1.4 shows the length of the exacerbation-free periods by treatment arm. This plot shows that each treatment arm has patients with very long time periods without an exacerbation (as mentioned earlier, we used a time cut-off of 3 years = 1095 days), but it appears that the waiting times to exacerbation are longer in the high dose group compared to the placebo group, and even for the low dose group the waiting times seem to be longer compared to the placebo group. This indicates that there might be a treatment effect. For the time in exacerbation (Figure 1.5), we see that each treatment group has quite extreme observations, with exacerbations as long as 300 days. The median exacerbation length for the high dose group and low dose group seem to be slightly lower compared to the placebo group.

Another way of summarizing these data is to use the proportion of time in exacerbation; this takes the dropout into account and gives only one number per individual (whereas there are usually several exacerbations per

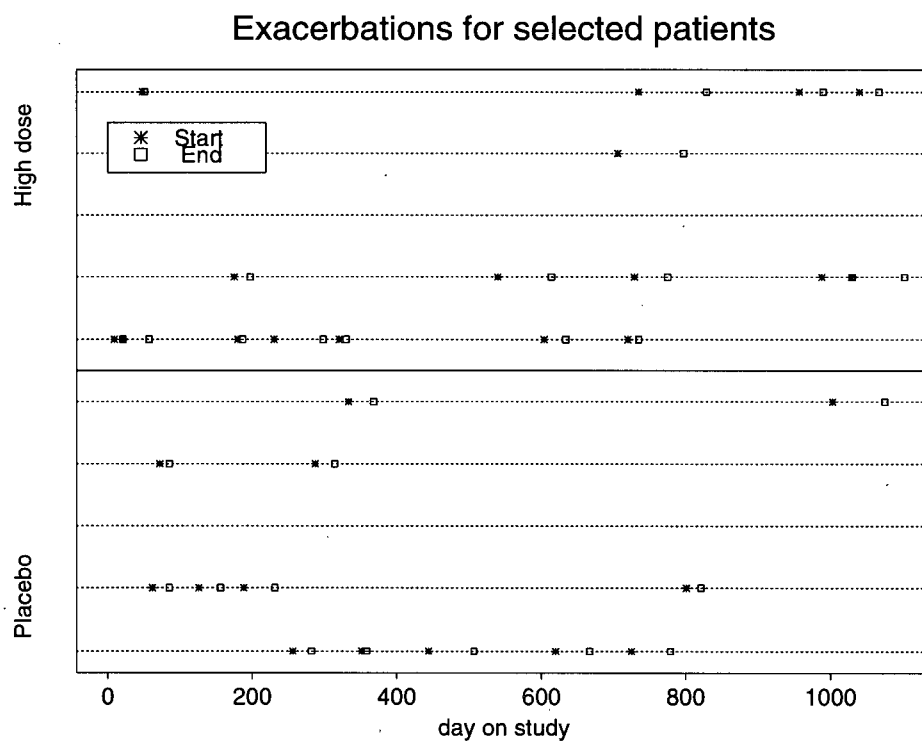


Figure 1.3: Patterns of exacerbations for selected patients

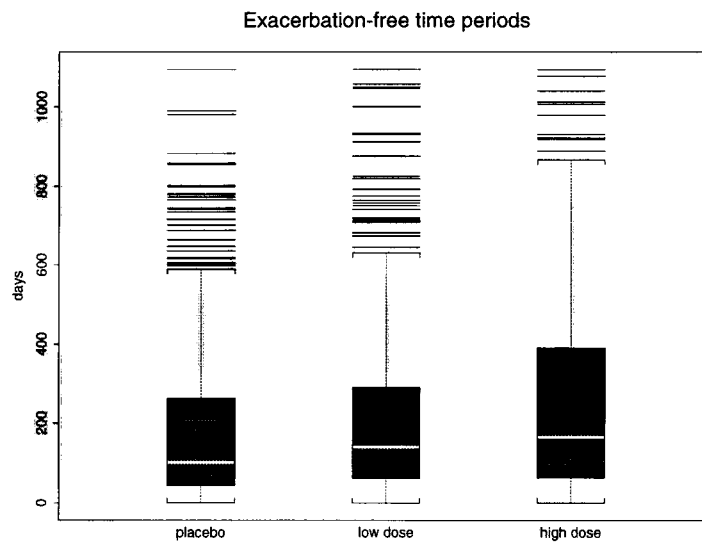


Figure 1.4: Length of exacerbation-free periods by treatment group

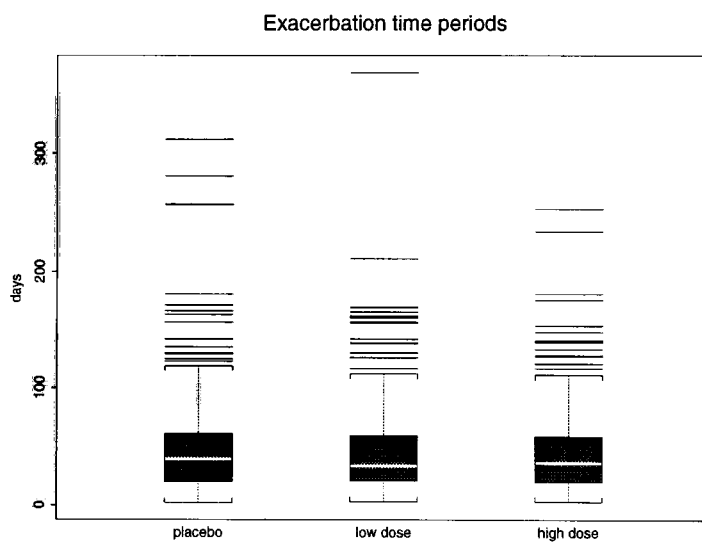


Figure 1.5: Length of exacerbations by treatment group

patient). We do not consider this outcome in any of the models in this thesis; we included it only for the purpose of data exploration since it gives a useful summary of the data on one individual that incorporates the length of exacerbations and allows us to investigate patterns over time and heterogeneity of the patients. In a disease process like MS, we expect that the exacerbation data early in the trial might differ from the exacerbation data late in the trial. To get an assessment of the pattern over time by treatment arm, we cut the time axis into intervals of 200 days (1-200, 201-400, 401-600, 601-800 and 801-1000 days) and computed the average proportion of time in exacerbation (number of days in exacerbation divided by number of days on trial) for each treatment arm; the results are shown in Figure 1.6. This plot shows that the treatment groups are separated quite consistently, with high dose having the lowest average proportion and placebo having the highest average proportion of time in exacerbation (except for one exception at time period 2, where the low dose and placebo arms have almost the same proportion). It also shows that the proportion of time in exacerbation decreases over time in the low dose and placebo groups; in the high dose group there is not much evidence of a decrease.

Figure 1.7 shows the proportion of time in exacerbation over the entire three years versus the covariates. Note that we have one proportion > 1 , because according to the data set this person left the trial before the recorded end of an exacerbation (so that the recorded time on study was less than the recorded time in exacerbation). This plot shows that the proportions are quite variable, ranging from 0 to 1. There is an indication that the proportion is higher in female patients than in male patients and patients with a lower time since onset tend to have a higher proportion of time in exacerbation. Age and EDSS do not have any apparent effect on the proportion of time in exacerbation. For the treatment groups, it seems that the proportion of time in exacerbation is higher in the placebo group compared to low dose and high dose groups, and the proportion seems to be a little higher in the low dose group compared to the high dose group.

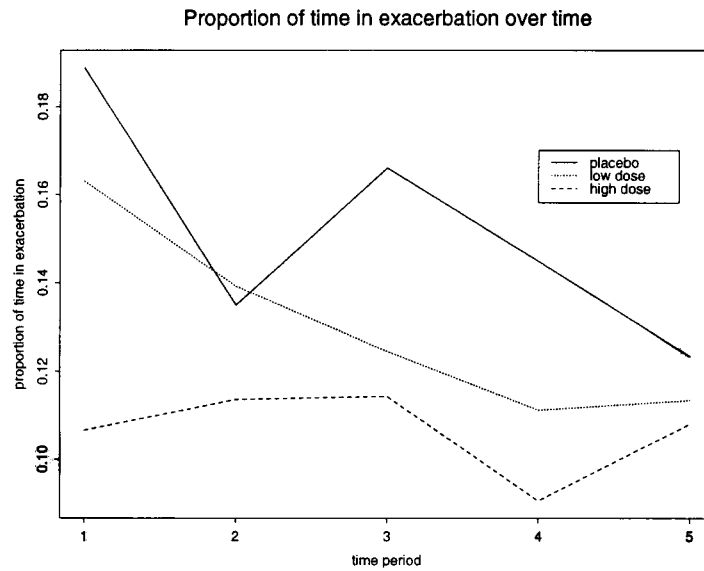


Figure 1.6: Proportion of time in exacerbation over time in 200 day periods by treatment

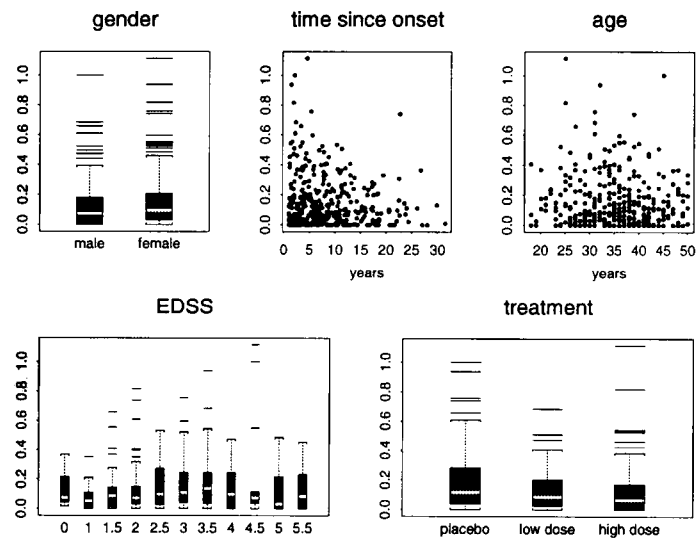


Figure 1.7: Proportion of time in exacerbation versus covariates

1.3 Previous Analyses and Our Objectives

As mentioned in the previous section, this data set has already been analyzed. The analyses presented for the exacerbation data in the paper describing the clinical results of the trial [9] focus on the annual exacerbation rates and the proportion of patients remaining exacerbation-free (the analysis in the other clinical trials was similar with respect to the exacerbation data). These data were also analyzed by D'yachkova [6]. She performed a number of different analyses for the exacerbation data, all based on the repeated binary responses for each patient obtained by cutting the time-axis into 6-week intervals and then recording if an exacerbation began in this interval.

These previous analyses use only the information on when an exacerbation started; they ignore the information available on the length of exacerbations. Figure 1.5 shows that the variability of the length of exacerbations is quite high. Ignoring this information means that we judge the efficacy of the treatment only by the number of exacerbations; a series of short exacerbations is considered worse for the patient than one long exacerbation. Given that some exacerbations are quite long, this is a questionable criterion for judging the treatment effect.

In contrast to these analyses, we will regard the data set as describing a two-state disease process. State 1 refers to the exacerbation-free state, and state 2 is the state where patients are in exacerbation. Our models will incorporate this second state and hence distinguish between exacerbations of different lengths. We investigate three basic stochastic models to analyze these data, which we introduce in the next section. The outcome variables we consider differ from model to model, but all incorporate some of the information available on the lengths of exacerbations. The aim of this thesis is to implement and compare several models that allow us to incorporate the information on the lengths of exacerbations. We did not attempt a comprehensive analysis of the data set with any of these models. Rather, the aim here is to see what can be achieved and which questions can be addressed with these different models.

There are two measures available in this data set to distinguish different

exacerbations: length and severity. We will only use the length to distinguish different exacerbations, although some of the models proposed could also be extended to incorporate the severity of exacerbations. In hindsight, the incorporation of severity could have been quite interesting, but we did not attempt this. We will discuss this point further in Chapter 6.

We are primarily interested in the assessment of the treatment effects on the outcomes related to the exacerbations data, but patterns in the data over time and the effect of the baseline covariates are also of interest. The distinction between baseline covariates and treatment groups is not sharp, and we will sometimes also view the treatments as covariates. The assessment of the effect of covariates can be done in two ways:

- assessment of the impact of covariates at the population level (i.e. on the distribution of the outcome variable in the patient population)
- assessment of the impact of covariates on a patient's disease process

Both these assessments are of interest. The first allows us, for example, to describe a treatment effect in a clearly interpretable way, so that we can quantify the beneficial effect of the treatment in clinically meaningful ways. The second assessment gives us more insight into the disease process to identify risk factors. The assumptions that are needed to address the first assessment might also be too restrictive, since we typically have to assume that the subgroups that are characterized by explanatory variables are relatively homogeneous. In a highly variable disease process like MS, this might not be appropriate.

For linear models with continuous data, regression coefficients from models that address the first assessment can be translated into regression coefficients that address the second assessment and vice versa. Regression coefficients from models for non-continuous responses that address these assessments usually have different interpretations, so that regression parameters on an individual level do not necessarily have an interpretation on the population level (see [7], p. 137).

1.4 Models Considered

We have seen in the previous section that different models for the analyses of the exacerbation data address different questions with regard to the assessment of covariates. Three types of stochastic models will form the basis of our analyses of the data:

- Markov chains
- alternating renewal processes
- marked Poisson processes

There are two general features we need to consider for these models: the stochastic aspects of the data set the model can capture, and the approach to estimation and inference for the model. In this section we describe the stochastic properties of these models and the advantages or disadvantages of each model and what questions concerning the covariates each model can address. The next section describes the approaches to inference for each model.

1.4.1 Markov Chains

Markov chains provide a convenient stochastic model for categorical longitudinal data. In our application, we have two categories (not in exacerbation, in exacerbation) and the Markov chain models the transition probabilities between these two states. This means we model the probability of the response being equal to category k at time j conditional on responses at times $l < j$ and the covariates. In contrast, for independent categorical data the probability of the response being equal to category k is modeled conditionally only on the covariates but not on previous responses.

This approach incorporates the longitudinal nature of the data by conditioning on previous responses. The dependence over time and the dependence on covariates is expressed in one model. Thus, it is a simple model for the dependence of categorical data over time and the model fitting is straightforward. The major disadvantage is that the interpretation of parameters for the covariates is in terms of the transition probabilities. This means that we have to interpret the regression parameters conditional on the previous responses.

This makes it easy to assess the impact of covariates on the disease process, but does not give a clear interpretation of their impact at the population level.

Another disadvantage is that the lengths of exacerbations enter the model only indirectly through the conditioning on previous responses. In this data set with fairly long exacerbation times, we need to consider a large number of previous responses if we use days as our unit of time since we expect the transition probability to decrease with the length of the current exacerbation.

1.4.2 Alternating Renewal Processes

We alluded in the previous section to the fact that the differences between the lengths of different exacerbations is difficult to capture in a Markov chain model. A model that takes the length of exacerbations directly into account is that of alternating renewal processes. Renewal processes are commonly used as models for repeated failure times and alternating renewal processes can be used to model a process that alternates between two (or more) states. In our application, we have two states where state 1 refers to the exacerbation-free state and state 2 to the in exacerbation state.

Renewal processes are defined in terms of their conditional intensity function. The intensity function at time t can be interpreted as the probability of an event occurring in the next instant given the history up to time t . The defining characteristic of a renewal process is that the model for this intensity incorporates the time since the last event (called the backward recurrence time).

To extend this model to a two-state process, we need to model the transition from state 1 to state 2 as well as the transition from state 2 to state 1. When both these transitions are modeled according to a renewal process (i.e. the transition probability is modeled conditionally on the time since entering this state) and an individual process alternates between these states, this model is called an alternating renewal process. We incorporate covariates into the parameterization for the conditional intensity functions. This means that, as was the case for the Markov chain model, we assess the impact of covariates on the transition probabilities.

The advantage of the alternating renewal process model over the Markov chain model is that the lengths of exacerbations are incorporated directly into the model through the backward recurrence time. The model has the shortcoming that we still assess covariates by their impact on the transition probabilities, which limits the interpretation of the results. An approach that models the mean counts and mean length of exacerbations directly is presented in the next section.

1.4.3 Marked Point Processes

The two previous models focus on the transition probabilities, conditional on either the response at the previous time-point or the time spent in the current state. This leads to mathematically convenient models, but the interpretation of the effects of covariates on the population level is not straightforward. To have directly interpretable results we want to model the mean counts of exacerbations and the mean length of exacerbations. This can be accomplished through marked Poisson processes and more generally, through marked point processes.

Poisson processes model the marginal means of the event counts at each time-point. They are popular models for repeated events partly because they are easy to fit and the interpretation of parameters is straightforward. Similar to renewal processes they can also be specified in terms of their intensities. They can be characterized by the fact that the inter-event times have exponential distributions. For a more detailed discussion, see Chapter 3.

Marked Poisson processes are a generalization of Poisson processes. In a Poisson process, the events have no associated properties; all events are identical. In marked processes, the events are distinguished by associated *marks*. These marks can be random variables; in our application they will count the number of days a patient was in exacerbation. Hence in this model events are not just points in time but have associated lengths.

Poisson processes assume that the counts of events occurring in disjoint time intervals are independent and have a Poisson distribution. Equivalent to

this assumption is that the inter-event times are independent and have an exponential distribution. These assumptions about the distribution of the counts and the inter-event times are too restrictive for our longitudinal data set, so we need to adopt a more general model within the very general class of *point processes*. To incorporate the lengths of exacerbations, we also introduce marks for this model which leads to a marked point process model. Marked point processes are generalizations of marked Poisson processes where the process for the event counts is a point process instead of a Poisson process. Chapter 3 gives more background on point processes.

1.5 Methods of Analysis

In the previous section we discussed how the different models incorporate the two states and the advantages or disadvantages of these models with respect to the interpretation of covariates. In this section we discuss how the analysis for these models proceeds. By this we mean:

- How does the exacerbation data enter into the analysis ?
- How do the analyses incorporate the longitudinal aspect of the data set ?

Before we proceed with the discussion for the different models, we first describe the coding for the baseline covariates, since this is the same for all models. The baseline covariates were coded as follows:

- treatment: we included the three arms as categorical variables, with the low dose and high dose included as indicator variables so that the treatment effects reflect the difference from placebo
- time since onset of MS: as a continuous variable in years
- EDSS: this is in fact an ordinal variable, but we treated it as a continuous covariate
- gender: female as 1, male as 0
- age: as a continuous variable in years

We used only time-independent covariates except for time on study itself. This simplifies the theoretical development and the computations, but some of the methods could also be applied with time-varying covariates.

All patients were exacerbation-free when they entered the trial, so one might want to take this into account when modeling the data. We did not consider this in our models.

For each modeling approach, we faced a considerable scope of possible models. However, the objective in this thesis is to provide a comparison of modeling approaches for a two-state disease process, not the comprehensive analysis of this particular data set. Hence we simplified this task and limited ourselves to just fitting the main effects. Further, we did not drop the non-significant covariates from the models; this also makes the comparison between the results obtained in different chapters easier.

1.5.1 Markov Chains

We have seen in Section 1.2 that most times in exacerbation and times not in exacerbation are relatively long (Figure 1.5). This implies, for example, that the transition probabilities out of exacerbation depend on how long this patient has been in exacerbation, since the probability an exacerbation ends is relatively low for the first few days but higher if the exacerbation has already lasted for a while. To capture this long-range dependence, we would need to incorporate a relatively large number of previous responses into the conditional specification (the number of previous responses is called the *order* of the chain). This introduces additional parameters into the model and makes the interpretation of covariates even more dependent on the disease process. To avoid having to consider high order chains, we collapsed the data over time. This is commonly done for data that has been collected over time to reduce the autocorrelation, although it results in some loss of information. To collapse the data, we considered only the data at every 10th day the patient was on the study and ignored all days in between. In this way we obtain a sequence of up to 110 data points, which is simply the subset consisting of the observation on every 10th day from the original daily sequence of 1095 days. Patients that drop out of the study before the end of the third year hence have a shorter

sequence of observed values.

The simplest versions of such models would assume common transition probabilities for individuals with the same covariates. In the case of MS, we have highly variable disease processes so this assumption might be too restrictive. By introducing random effects into the model, we need only assume that the process is a Markov chain conditionally on these random effects, but not at the population level.

1.5.2 Alternating Renewal Processes

The alternating renewal process model incorporates the exacerbation data as the lengths of time intervals in state 1 (not in exacerbation) or state 2 (in exacerbation). Hence we can use the original data without collapsing it and incorporate the length of time intervals in days. For intervals not observed completely because the patient drops out before the next transition, we model the survival function for these censored times. This is analogous to the common approach in survival analysis.

A simple model is the fixed effects model which assumes that the intensities are determined by the covariates and the backward recurrence time. This model assumes that for both states the sojourn times for an individual are independent random variables. To allow for correlated sojourn times on one individual and to accommodate heterogeneity between the individuals, we introduce random effects into the model.

1.5.3 Marked Point Processes

As mentioned in Section 1.4.3, our marked point process approach models the mean event counts and the associated sum of the length of exacerbations at each time-point. Although an extension of this model to a continuous time setting is possible, we only consider a fixed number of time-points. That is, we choose a fixed set of time-points and then count the number of exacerbations and the sum of the lengths of exacerbations in the intervals between these time-points. Hence each patient contributes a sequence of pairs of numbers: the number of exacerbations starting in each time-period and the sum of the

lengths of these exacerbations for each time-period. If a patient drops out before the end of the current interval, data for this and all subsequent intervals are missing.

We have mentioned previously that Poisson processes assume independence of the data within individuals, and thus are not suitable as models for longitudinal data where we expect the counts on an individual to be correlated over time. However, more complicated models based on the conditional intensity of the process lose the simplicity of interpretation of parameters. To overcome this difficulty, we use a GEE approach in a point process setting. This approach allows us to model the regression of the response on the covariates at each time-point separately from the correlation within individuals over time. This means we can use the estimating equations for a Poisson process model to obtain estimates for the regression parameters and then obtain standard errors that are valid more generally by using the so-called sandwich estimator for the covariance matrix. This inference is valid for any point process that has the same marginal mean event counts, no matter what the correlation between these counts. This is referred to as the robustness property of this approach.

1.6 Outline of this Thesis

The remainder of this thesis is organized as follows: Chapter 2 discusses the Markov chain model; Chapters 4 and 5 deal with the implementation of the alternating renewal processes and the marked point processes model, respectively. Since both Chapters 4 and 5 use point processes as their underlying framework, we introduce some general properties in Chapter 3. This chapter also includes the initial data analysis for Chapters 4 and 5, since in both cases it relies mainly on Poisson processes. We conclude with a comparison of the approaches and some general discussion in Chapter 6.

Chapter 2

A Markov-Chain Model

2.1 Introduction

In this chapter, we model each individual patient's disease process as a Markov chain. Each patient is represented by a sequence of zeroes and ones corresponding to the states "no exacerbation" or "exacerbation" (we explain the definition of the response in more detail below). We model the transition probabilities between these two states.

In Chapter 1 we mentioned that the model for the transition probability at time j can involve any fixed number (which is called the order of the chain) of responses y_l at previous time points $l < j$. Figure 1.5 shows that the exacerbations usually last at least a few days, hence we would need to consider a relatively high order for the chain if we modeled the daily transitions to capture this long-range dependence. This would introduce more parameters into the model and also make the interpretation of the covariates more dependent on the disease process. To overcome this difficulty, we collapse the data on each individual. More precisely, we fix a number of time points in advance and record for each patient if on this day the patient is in exacerbation (recorded as 1; referred to as state 2) or not (recorded as 0; referred to as state 1). This is done for all of the time points where the patient was still on the trial (i.e. the dropout occurred after this time point). This way we have common time points on which all patients are observed, but sequences of different length for different patients.

We develop parametric models for the transition probabilities between these states at successive time points. We first consider a fixed effects model which assumes common transition probabilities for individuals with the same covariates. To allow for different underlying disease activity for different individuals that is not explained by covariates, subsequently we introduce random effects into this model.

The body of literature on Markov chains from a mathematical perspective is vast. Papers dealing with statistical inference for Markov chains are Anderson and Goodman [2] and Kalbfleisch and Lawless [13]. Random effects are commonly used in models for repeated observations, see Lawless [16] or Diggle, Zeger and Liang [7], for example. For applications of random effects models to Markov chains, see Stiratelli, Laird and Ware [30] or Cook and Ng [3].

We first introduce some general theory for inference on Markov chain models and introduce a parametric model for the transition probabilities. We then extend this fixed effects model to a random effects model. We apply these methods to the MS data set and conclude with a goodness-of-fit assessment and some discussion of the results.

2.2 The Likelihood

In this section we introduce the notation and derive the likelihood for general two-state Markov chains. We assume the response Y_{ij} for individual i at time j is binary and that the distribution of the response Y_{ij} given the response $Y_{i,j-1}$ is independent of all earlier responses $(Y_{i,j-2}, Y_{i,j-3}, \dots)$. This is called a first-order Markov chain. Hence we model $P(Y_{ij} = y_{ij} | Y_{i,j-1} = y_{i,j-1}, \mathbf{x}_{ij})$; that is, the response at time j given the response at time $j - 1$.

We assume that we observe individual i , $i = 1, \dots, m$, at a set of equally spaced times $j = 1, \dots, n_i$. This is convenient since then we don't have to account for time intervals of different lengths. Let us denote the vector of observed responses $y_{ij} \in \{0, 1\}$ for this individual by \mathbf{y}_i . Then the probability

of this realization \mathbf{y}_i for the random sequence \mathbf{Y}_i becomes, conditional on Y_{i1} ,

$$P(\mathbf{Y}_i = \mathbf{y}_i) = \prod_{j=2}^{n_i} P(Y_{ij} = y_{ij} | Y_{i,j-1} = y_{i,j-1})$$

by virtue of the Markov property.

Since Y_{ij} is a Bernoulli random variable, we have

$$P(Y_{ij} = 1 | Y_{i,j-1} = y_{i,j-1}) + P(Y_{ij} = 0 | Y_{i,j-1} = y_{i,j-1}) = 1. \quad (2.1)$$

Together with the assumption that sequences from different individuals are independent, this leads to the likelihood, conditional on Y_{i1} , of:

$$L = \prod_{i=1}^m \prod_{j=1}^{n_i} P(Y_{ij} = 1 | Y_{i,j-1} = y_{i,j-1})^{y_{ij}} (1 - P(Y_{ij} = 1 | Y_{i,j-1} = y_{i,j-1}))^{1-y_{ij}}. \quad (2.2)$$

This likelihood is very similar to a likelihood for independent Bernoulli random variables. This fact is exploited below to obtain estimates.

For an initial analysis ignoring the baseline covariates, we are interested in nonparametric estimates of the transition probabilities. To allow us to explore the changes over time separately for each treatment group, we cut the time-axis into intervals and examine the data stratified by these time intervals and by treatment group. Hence we get an estimated transition matrix for each treatment group and time interval.

Let us denote the transition probabilities (which are the parameters here) by

$$p_{gh} = P(Y_{ij} = h | Y_{i,j-1} = g)$$

for $g, h = 0, 1$ where individual i is in a specified treatment group and time j is in a specified time interval. For notational convenience we do not include the treatment group and time-interval in the notation. For each treatment group and time interval, the following likelihood for the parameters p_{gh} , $g, h = 0, 1$ can be obtained using expression (2.2)

$$\tilde{L} = \prod_g \prod_h p_{gh}^{n_{gh}},$$

where n_{gh} is the number of g to h transitions for this treatment group in this time interval. Since the constraints are given by (2.1), \tilde{L} is proportional to a binomial density. Hence the MLEs in this case are the same as for a binomial distribution, namely

$$\hat{p}_{gh} = \frac{n_{gh}}{\sum_h n_{gh}}.$$

2.3 A Parametric Fixed Effects Model

As we have seen in the previous section, we can parametrize the likelihood in terms of the transition probabilities $P(Y_{ij} = 1 | Y_{i,j-1} = y_{i,j-1})$. To incorporate the covariates into the model, we model the dependence of the transition probabilities $P(Y_{ij} = 1 | Y_{i,j-1} = y_{i,j-1})$ on the vector of covariates \mathbf{x}_{ij} for individual i at time j as

$$\text{logit } P(Y_{ij} = 1 | Y_{i,j-1} = 0, \mathbf{x}_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta}_0, \quad (2.3)$$

$$\text{logit } P(Y_{ij} = 1 | Y_{i,j-1} = 1, \mathbf{x}_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta}_1. \quad (2.4)$$

Here $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ are vectors of regression parameters and we assume that each model contains an intercept term. This way, we obtain separate regression parameters for the different transitions. $\boldsymbol{\beta}_0$ is the parameter vector for the probability to be in exacerbation when the patient was not in exacerbation at the previous time point, and $\boldsymbol{\beta}_1$ is the parameter vector for the probability to be in exacerbation when the patient was in exacerbation at the previous time point.

The parametrizations (2.3) and (2.4) make a combined inference for both states more difficult, since they are stratified by the response at the previous time point. To allow for more general models, we follow Diggle, Liang and Zeger [7] and parametrize this model in a more concise way. To judge if the effects of covariates are the same for both transitions in the first-order Markov chain model, we re-express (2.3) and (2.4) as

$$\text{logit } P(Y_{ij} = 1 | Y_{i,j-1} = y_{i,j-1}, \mathbf{x}_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + y_{i,j-1} \mathbf{x}_{ij}^T \boldsymbol{\alpha} \quad (2.5)$$

so that $\boldsymbol{\beta}_0 = \boldsymbol{\beta}$ and $\boldsymbol{\beta}_1 = \boldsymbol{\beta} + \boldsymbol{\alpha}$. If an element of $\boldsymbol{\alpha}$ is estimated as being significantly different from 0, this indicates that the effect of the covariate on

the probability of being in exacerbation depends on whether the patient was in exacerbation at the previous time point or not.

We use a similar parametrization to compare the first-order model to a second-order model. We do not attempt to find the most appropriate order of the chain; we simply want to assess the adequacy of the first-order model compared to a second-order model and investigate departures from the assumption of order 1. For a chain of order 2, the transition probabilities depend on the responses at the previous two time points. Analogously to (2.3) and (2.4) we can fit 4 separate logistic regressions with parameter vectors $\beta_{00}, \beta_{01}, \beta_{10}$ and β_{11} for the history $(Y_{ij-2}, Y_{ij-1}) = (0, 0), (0, 1), (1, 0)$ and $(1, 1)$, respectively. We can express these 4 models in a single regression:

$$\text{logit } P(Y_{ij} = 1 | Y_{i,j-1} = y_{i,j-1}, Y_{i,j-2} = y_{i,j-2}, \mathbf{x}_{ij}) = \mathbf{x}_{ij}^T \beta + y_{i,j-1} \mathbf{x}_{ij}^T \alpha_1 + y_{i,j-2} \mathbf{x}_{ij}^T \alpha_2 + y_{i,j-1} y_{i,j-2} \mathbf{x}_{ij}^T \alpha_3. \quad (2.6)$$

Hence $\beta_{00} = \beta, \beta_{01} = \beta + \alpha_1, \beta_{10} = \beta + \alpha_2$ and $\beta_{11} = \beta + \alpha_1 + \alpha_2 + \alpha_3$. If the Markov chain is in fact of order one, the transition probability does not depend on the outcome Y_{ij-2} and thus $\beta_{00} = \beta_{10}$ and $\beta_{01} = \beta_{11}$ or, in other words, $\alpha_2 = 0$ and $\alpha_3 = 0$. So we can use a likelihood-ratio test of $H_0 : \alpha_2 = 0, \alpha_3 = 0$ to assess if a chain of order 2 would significantly improve the fit. If this is the case, we can examine the nature of the departure from the first-order model in by examining the fit of model (2.6).

Note that for estimation purposes this model treats the previous responses as a covariate. Hence we can utilize GLM-software to fit this model.

2.4 A Random Effects Model

In this section we extend the fixed effects model given by (2.3) and (2.4) from the previous section to a random effects model. The likelihood for the Markov chain model in the previous section resembles the likelihood for independent Bernoulli variables. We can exploit this connection further by using MIXOR, software for fitting mixed effects ordinal regression models, to fit a random effects model. We now describe the software and how we utilize it.

MIXOR is a program for mixed effects ordinal regression models developed by Hedeker and Gibbons (MIXOR: A Computer Program for Mixed-effects Ordinal Probit and Logistic Regression Analysis, available on <http://www.uic.edu/~hedeker>). For simplicity and since this is our case, we limit our discussion to the case of binary responses with a logistic link.

The basic idea underlying the model is that the binary outcome is reflecting an unobservable continuous latent variable. The two variables are related by the so-called threshold concept: a response occurs when the latent variable exceeds the threshold value. The random effects are incorporated on the latent variable scale.

We denote the latent response by Z_{ij} and the observed binary variable by Y_{ij} for individual i at time j . We model the latent response for individual i as

$$\mathbf{Z}_i = \mathbf{X}_i \boldsymbol{\nu} + \mathbf{W}_i \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i,$$

where \mathbf{X}_i is the design matrix for the fixed effects and \mathbf{W}_i is the design matrix for the random effects. The errors ϵ_{ij} are assumed to be independent random variables following a logistic distribution with mean 0 and common variance. The random effects $\boldsymbol{\eta}_i$ are assumed to be independent of the errors and to be multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$.

The log odds of the event $Y_{ij} = 1$ for a given individual i are then modeled as

$$\text{logit } P(Y_{ij} = 1 | \boldsymbol{\nu}, \boldsymbol{\eta}_i) = \tilde{z}_{ij}, \quad (2.7)$$

where $\tilde{z}_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\nu} + \mathbf{w}_{ij}^T \boldsymbol{\eta}_i$. Alternatively, we can write

$$P(Y_{ij} = 1 | \boldsymbol{\nu}, \boldsymbol{\eta}_i) = \Psi(\tilde{z}_{ij}),$$

where $\Psi(u) = \exp(u)(1 + \exp(u))^{-1}$.

With these model assumptions, the contribution to the likelihood for individual i becomes

$$L_i(\mathbf{y}_i | \boldsymbol{\nu}, \boldsymbol{\eta}_i) = \prod_{j=1}^{n_i} P(Y_{ij} = 1 | \boldsymbol{\nu}, \boldsymbol{\eta}_i)^{y_{ij}} (1 - P(Y_{ij} = 1 | \boldsymbol{\nu}, \boldsymbol{\eta}_i))^{1-y_{ij}}.$$

The marginal likelihood for individual i is then given by

$$h(\mathbf{y}_i) = \int L_i(\mathbf{y}_i | \boldsymbol{\nu}, \boldsymbol{\eta}) g(\boldsymbol{\eta}) d\boldsymbol{\eta},$$

where g is the multivariate normal distribution of the random effects.

MIXOR uses Fisher's scoring algorithm to get maximum marginal likelihood estimates for the parameters, which involves numerical integration to solve the maximum marginal likelihood equations. The numerical integration is implemented through Gauss-Hermite integration (see Section 4.6 for a description of this method). The information matrix at convergence is then used to get the large sample variances and covariances.

So far we have modeled Bernoulli random variables Y_{ij} with a random effect, that is we have modeled $P(Y_{ij} = 1 | \boldsymbol{\nu}, \boldsymbol{\eta}_i, \mathbf{x}_{ij})$. To make the connection to the Markov chain model, we now model the transition probabilities conditional on the random effects; that is, we now model the probability of being in exacerbation conditional on the random effects, the covariate vector *and* the response at the previous time point. We denote the random effects for the transition into exacerbation by $\boldsymbol{\gamma}_i$ and the random effects for the exacerbation to exacerbation transition by $\boldsymbol{\delta}_i$. In analogy to (2.3) and (2.4), we can obtain regressions separately for each state as

$$\text{logit } P(Y_{ij} = 1 | Y_{i,j-1} = 0, \mathbf{x}_{ij}, \boldsymbol{\gamma}_i) = \mathbf{x}_{ij}^T \boldsymbol{\beta}_0 + \mathbf{w}_{ij}^T \boldsymbol{\gamma}_i, \quad (2.8)$$

$$\text{logit } P(Y_{ij} = 1 | Y_{i,j-1} = 1, \mathbf{x}_{ij}, \boldsymbol{\delta}_i) = \mathbf{x}_{ij}^T \boldsymbol{\beta}_1 + \mathbf{w}_{ij}^T \boldsymbol{\delta}_i. \quad (2.9)$$

2.5 Model Assessment

To assess the appropriateness of the fixed and random effects models we follow the approach of Muenz and Rubinstein [21]. They compare the estimated occupation probabilities with the observed frequencies for each state. We first discuss this in the context of fixed effects and then extend it to the context of random effects.

A single Markov chain with two states switches between these states over time. The occupation probability of the chain at time j is determined by

the response at the first time-point and the transition probabilities for times $l < j$, since

$$\begin{aligned} P(Y_{ij} = y_{ij}) \\ = P(Y_{ij} = y_{ij} | Y_{i,j-1} = 0)P(Y_{i,j-1} = 0) + P(Y_{ij} = y_{ij} | Y_{i,j-1} = 1)P(Y_{i,j-1} = 1) \end{aligned}$$

and

$$P(Y_{i2} = y_{i2}) = P(Y_{i2} = y_{i2} | Y_{i1} = 0)(1 - Y_{i1}) + P(Y_{i2} = y_{i2} | Y_{i1} = 1)Y_{i1}.$$

Applying these equations recursively, we obtain

$$\begin{aligned} & \begin{pmatrix} P(Y_{ij} = 0) \\ P(Y_{ij} = 1) \end{pmatrix} \\ &= \left\{ \prod_{j=2}^t \begin{pmatrix} P(Y_{ij} = 0 | Y_{i,j-1} = 0) & P(Y_{ij} = 0 | Y_{i,j-1} = 1) \\ P(Y_{ij} = 1 | Y_{i,j-1} = 0) & P(Y_{ij} = 1 | Y_{i,j-1} = 1) \end{pmatrix} \right\} \begin{pmatrix} 1 - Y_{i1} \\ Y_{i1} \end{pmatrix}. \end{aligned}$$

To get estimates $\hat{P}(Y_{i,t} = 0)$ and $\hat{P}(Y_{i,t} = 1)$ for the fixed effects model given by (2.3) and (2.4), we simply plug in $\hat{\beta}_0$ and $\hat{\beta}_1$ in the respective parametrizations. For the random effects model, we also need to estimate the individual random effects. This is discussed at the end of this section.

We want to compare the estimated occupation probabilities to the observed relative frequencies $F(k, j)$ for subsets of individuals, given by

$$F(k, j) = \frac{|\{i : Y_{ij} = k, i \in G\}|}{|G|}$$

for $k = 0, 1$ and G a subset of $\{1, \dots, m\}$. G will be determined by the quartiles of the estimated transition probabilities. We could also take G to be the set of all individuals, but by using (for example) 4 subsets we get a more refined assessment to detect potential systematic differences between these subsets.

We evaluate the estimated exacerbation to exacerbation transition probabilities at one point in time (we arbitrarily took day 500) and group individuals according to the quartile of their estimated no exacerbation to exacerbation

transition probability on that day. For each quartile, we calculate the observed occupation frequencies at each time point and the average occupation probabilities at each time point and plot these two time series to graphically assess the lack of fit.

To apply the same method to the random effects model, we need to estimate the transition probabilities which are modeled conditionally on the unobservable random effects. We want to estimate the individual random effects γ_i and δ_i and plug these estimates into the conditional specification to obtain estimates for the transition probabilities. This procedure is described in the next paragraph.

We have specified a prior distribution for the random effects given by the density $g(\gamma, \delta)$ and we have a sampling distribution for individual i given by

$$\begin{aligned} L_i(\mathbf{y}_i | \mathbf{x}_i, \beta, \alpha, \gamma_i, \delta_i) \\ = \prod_{j=2}^{n_i} P(Y_{ij} = 1 | \mathbf{x}_i, \beta, \alpha, \gamma_i, \delta_i)^{y_{ij}} (1 - P(Y_{ij} = 1 | \mathbf{x}_i, \beta, \alpha, \gamma_i, \delta_i))^{1-y_{ij}}. \end{aligned}$$

Hence the posterior distribution for γ_i and δ_i , given the data and α and β , is proportional to

$$L_i(\mathbf{y}_i | \mathbf{x}_i, \beta, \alpha, \gamma_i, \delta_i) g(\gamma_i, \delta_i)$$

which can be estimated as

$$L_i(\mathbf{y}_i | \mathbf{x}_i, \hat{\beta}, \hat{\alpha}, \gamma_i, \delta_i) g(\gamma_i, \delta_i).$$

To estimate the random effects, we use the means with respect to the posterior distribution with estimates for α and β and the variance for the random-effects plugged in.

2.6 Application to the Exacerbation Data

To achieve a reasonably low order for the Markov chain we collapsed the data as already mentioned by using only every 10th day. Since most of the exacerbations are a lot longer than 10 days (see Figure 1.5), we still observe most

exacerbations that occurred. We defined $y_{ij} = 1$, $j = 1, \dots, 110$, if patient i had an exacerbation on day j times 10, and $y_{ij} = 0$ otherwise.

To get an initial assessment for this data set, we split the time axis into 5 intervals (1-200, 201-400, 401-600, 601-800 and 801-1000 days) and obtained non-parametric estimates for the 10-day transition probabilities as described earlier for each interval and treatment group separately. The estimates for the no exacerbation to exacerbation transition given in Figure 2.1 indicate a treatment effect for the comparison of the high dose and placebo groups; there is not much difference between the low dose and placebo groups. There also seems to be a time-trend in all treatment groups with different trends for the treatment groups for the first 4 time-intervals and a drop-off in all groups in the last interval. The fact that the placebo group also seems to have a trend to have less exacerbations over time is consistent with findings in [6].

For the exacerbation to exacerbation transition (Figure 2.2) there seems to be a treatment effect only in the last time-interval, and overall there is no consistency in this effect over time. There also seems to be a time-effect which is similar in the low dose and placebo groups except for the last interval. In summary both plots show the need to incorporate time as a covariate and we anticipate a much more prominent treatment effect for the transition into exacerbation. A linear time trend seems a bit questionable, but not unreasonable. We explore this with parametric models in the next section.

We now implement the parametric models as outlined in the previous sections. We have already stated in Chapter 1 that we use the data set to motivate our models, but a comprehensive analysis is not our goal. This should be kept in mind for this application of the Markov chain model.

2.6.1 Fixed Effects Models

We first fit a first-order model with the parsimonious parametrization (2.5). This model allows us to judge if the effect of the covariates is different if the patient was in exacerbation at the previous time point or not. Table 2.1 shows the estimated effects. A significant value for a component of α means the effect of that covariate is different for the different transitions. We see that except for the time since onset, EDSS and the low dose treatment, the effects

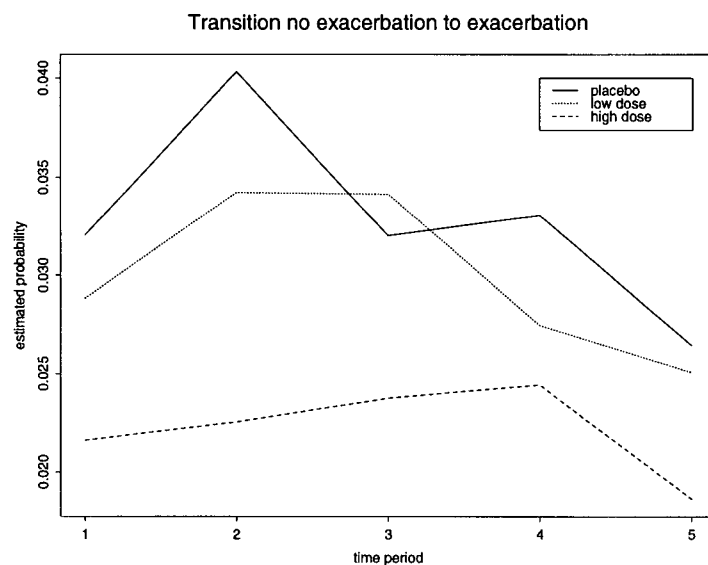


Figure 2.1: No exacerbation to exacerbation transition probabilities by time period

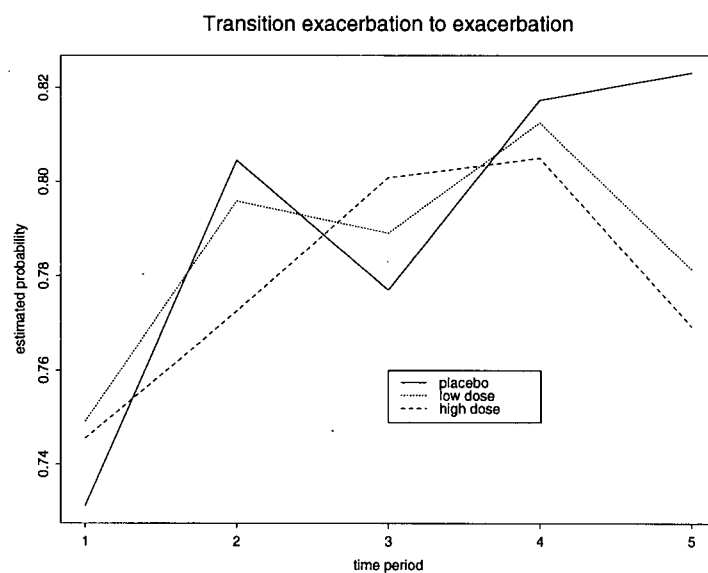


Figure 2.2: Exacerbation to exacerbation transition probabilities by time period

	β		α	
covariate	estimate	z-score	estimate	z-score
low dose	-0.1754	-2.20	0.0940	0.78
high dose	-0.4468	-5.25	0.4078	3.14
age	-0.0191	-3.43	0.0252	3.06
gender	0.2011	2.62	-0.2283	-1.94
time since onset	-0.0035	-0.55	-0.0063	-0.63
EDSS	0.0889	3.32	-0.0577	-1.37
time	-0.0051	-4.62	0.0068	4.00

Table 2.1: Covariate effects for the first-order fixed effects model (2.5)

	into exacerbation			exacerbation to exacerbation		
covariate	estimate	s.e.	z-score	estimate	s.e.	z-score
low dose	-0.1754	0.0798	-2.20	-0.0814	0.0904	-0.90
high dose	-0.4468	0.0850	-5.25	-0.0389	0.0979	-0.40
age	-0.0191	0.0056	-3.43	0.0061	0.0061	1.01
gender	0.2011	0.0766	2.62	-0.0273	0.0892	-0.31
time since onset	-0.0035	0.0064	-0.55	-0.0098	0.0075	-1.30
EDSS	0.0889	0.0268	3.32	0.0313	0.0323	0.97
time	-0.0051	0.0011	-4.62	0.0018	0.0013	1.34

Table 2.2: Covariate effects for the first-order fixed effects model, separate regressions for the transitions

of the covariates are in fact estimated to be different. Hence we re-express this model with the separate logistic regressions given by (2.3) and (2.4), since this gives more interpretable coefficients.

Table 2.2 shows the results for the separate regressions. For the no exacerbation to exacerbation transition, we see a strong beneficial effect of the high dose treatment (which we anticipated from the exploratory analysis), but there is also a significant treatment effect for the low dose treatment. There is a significant decrease in the transition probability over time, which we also expected after the exploratory analysis. Older patients are less likely to start an exacerbation as well as patients with a greater time since onset. Female

patients are also more likely to start an exacerbation, which indicates that gender has an impact not only on the prevalence but also on the disease process. Patients with a higher baseline EDSS are also more likely to have an exacerbation, which indicates a relationship between the disease progression measured by exacerbations and by EDSS. For the exacerbation to exacerbation transition, we find that no covariate is statistically significant. The treatments seem to be slightly beneficial, but the effects are very weak. Time has a modestly strong effect; patients are more likely to remain in exacerbation later in the trial. Time since onset also has a modest effect, with patients who have a greater time since onset being less likely to remain in exacerbation. We have seen previously that the effects of covariates are estimated to be different for the two transitions. This table shows that for most covariates this difference is significant because the effect is significant for the transition into exacerbation, but not significant for the exacerbation to exacerbation transition. One exception is age, where older patients have a significantly lower probability of starting an exacerbation but there is some indication that older patients have a higher probability of staying in exacerbation. The other exception is time; the probability of starting an exacerbation decreases over time but the probability of remaining in exacerbation increases over time.

We have previously mentioned that a first-order Markov chain might not be sufficient to capture the longitudinal structure of this data set and that it might require a higher order chain to capture this structure. We employ a different model in Chapter 4 that deals directly with the lengths of exacerbations, so we do not explore which order of the chain is most appropriate. We simply use the likelihood ratio test to test if a second-order model provides a better fit than the first-order model. The χ^2 test-statistic is 44.04 on 16 degrees of freedom ($p < 0.001$), hence the second-order model provides a significantly better fit. To examine the departures from the first-order model more closely, we fit the second-order model given by (2.6) and see how far we can reduce it to a first-order model.

The z-scores for model (2.6) with all baseline covariates included are shown in Table 2.3. If a first-order model was adequate, the effects of α_2 and α_3 would be weak. However, we see that there are some fairly strong estimated effects for these parameters. To simplify this model and to see how much we

	β	α_1	α_2	α_3
covariate	z-score	z-score	z-score	z-score
low dose	-2.21	0.12	0.40	-0.20
high dose	-5.28	1.90	1.04	-1.12
age	-3.53	1.03	0.98	-0.49
gender	2.57	-0.60	0.05	-0.25
time since onset	-0.20	0.86	-2.02	1.14
EDSS	3.07	0.86	1.19	-1.85
time	-4.65	2.99	0.68	-1.09

Table 2.3: Z-scores for covariate effects for the second-order fixed effects model (see (2.6)), initial model

can reduce it towards a first-order model, we used a backward selection (with a cutoff of 1.30 or -1.30 for the z-score) to drop components of α_2 or α_3 from the model. The estimates for the reduced model are shown in Table 2.4. This table shows that for most covariates, we can indeed eliminate the effects corresponding to interactions with the response two time-points previously. For EDSS and time since onset however, these interactions are quite strong, so that the interpretation of the effects for these covariates in the first-order model may not be quite appropriate.

We re-expressed these covariate effects in terms of the parameters $\beta_{00}, \beta_{10}, \beta_{01}$ and β_{11} to see the differences in the effects for these two-stage transitions; the results are shown in Table 2.5. To compare to the first-order model, we have indicated in the top row which transition these estimated effects correspond to in the first-order model ($y(t-1)$ represents the response at the previous time-point, so the two columns with the heading $y(t-1) = 1$ correspond to the exacerbation to exacerbation transition in the first-order model). In a first-order model, we have $\beta_{00} = \beta_{10}$ and $\beta_{01} = \beta_{11}$. We have seen already that for the covariates in this table, this is not the case; this table shows that for both cases ($y(t-1) = 1$ and $y(t-1) = 0$), the estimates differ considerably. This means that the interpretation of the covariate effects from the first-order model has to be viewed with some caution.

	β		α_1	
covariate	estimate	z-score	estimate	z-score
low dose	-0.1746	-2.18	0.0892	0.73
high dose	-0.4421	-5.19	0.3947	3.03
age	-0.0189	-3.40	0.0249	3.02
gender	0.2007	2.61	-0.2277	-1.92
time since onset	-0.0025	-0.38	0.0231	1.31
EDSS	0.0845	3.13	0.0500	0.64
time	-0.0050	-4.60	0.0072	4.20
	α_2		α_3	
covariate	estimate	z-score	estimate	z-score
low dose	—	—	—	—
high dose	—	—	—	—
age	—	—	—	—
gender	—	—	—	—
time since onset	-0.0369	-2.12	—	—
EDSS	0.1219	1.44	-0.2415	-3.71
time	—	—	—	—

Table 2.4: Covariate effects for the reduced (after backward selection) second-order fixed effects model (2.6), final model; a — indicates this term was dropped from the model

	$y(t-1) = 0$		$y(t-1) = 1$	
	β_{00}	β_{10}	β_{01}	β_{11}
covariate	estimate	estimate	estimate	estimate
time since onset	-0.0025	-0.0394	0.0206	-0.0163
EDSS	0.0845	0.2064	0.1345	0.0149

Table 2.5: Estimates for the second-order transition probabilities from the reduced second-order model (other estimates omitted)

	into exacerbation			exacerbation to exacerbation		
covariate	estimate	s.e.	z-score	estimate	s.e.	z-score
low dose	-0.2800	0.1454	-1.93	-0.0923	0.1041	-0.89
high dose	-0.5118	0.1435	-3.57	-0.0311	0.1303	-0.24
age	-0.0196	0.0102	-1.93	0.0052	0.0064	0.80
gender	0.2219	0.1292	1.72	-0.0100	0.1070	-0.09
time since onset	-0.0061	0.0111	-0.55	-0.0069	0.0093	-0.74
EDSS	0.0960	0.0518	1.86	0.0347	0.0377	0.92
time	-0.0037	0.0011	-3.34	0.0015	0.0017	0.90
σ	0.84	0.06		0.33	0.07	

Table 2.6: Covariate effects for the random intercept models (2.8) and (2.9). σ denotes the standard deviation for the random intercepts

2.6.2 Random Effects Models

In Chapter 1 we alluded to the fact that MS is a highly variable disease process. For this reason, we want to introduce random intercepts into the model to allow patients to have different tendencies to remain in exacerbation or start an exacerbation. A higher intercept for the transition into exacerbation indicates that this patient has a higher probability to start an exacerbation than another patient with the same covariates and a lower intercept. So this random intercept can explain differences between patients who have the same covariates, but different tendencies to start an exacerbation. It does not explain variabilities within patients, however. The random intercepts for the exacerbation to exacerbation transition have an analogous interpretation. Table 2.6 shows the results for the separate regressions for each state with random intercepts given by the model (2.8) and (2.9).

The results for the covariates are qualitatively similar to the results for the fixed effects model (although the interpretation here is conditional on the random intercepts). Again most covariates are significant for the transition into exacerbation, whereas no covariate is significant for the exacerbation to exacerbation transition. We will compare the fixed and random effects models in the next section. The estimated standard deviation for the random intercepts for the transition into exacerbation is fairly high, which means that the intercepts can explain some of the patient to patient variability of the

transition probability that is not explained by the covariates. The estimated standard deviation for the random intercepts for the exacerbation to exacerbation transition is also different from zero, although much smaller than the standard deviation for the transition into exacerbation.

We also tried to fit models with a random coefficient for the time effect for each state, however MIXOR did not converge for these models with these two random effects included.

2.7 Discussion

In this section we compare the results for the different first-order models and perform a goodness-of-fit assessment for the fixed effects model.

We want to compare the implications of these different fitted models, although as pointed out in Chapter 1, these coefficients have different interpretations in random effects and fixed effects models. However, one can still compare which covariates are the most important in the respective models and which lead to significant decreases or increases in the respective transition probabilities.

Although the interpretation of covariates is different in random and fixed effects models, there are some results available on the relationship of the regression coefficients from these models. Diggle, Liang and Zeger ([7], p. 140) discuss this for a logistic regression example (slightly different from our models). They show that in their situation, for any specific covariate, the ratio of the regression parameters from the fixed effects model and the corresponding random intercept model is a constant less than 1 that is monotonically decreasing in the variance of the random intercepts. The model considered here is not the same, but it is still interesting to examine these ratios to see if this ratio is consistent across different covariates. Table 2.7 shows the estimated effects for the fixed effects model given by (2.3) and (2.4) and the random intercepts model given by (2.8) and (2.9) as well as the ratios of the estimated effects. For the transition into exacerbation, the table shows that except for time the ratio is indeed smaller than 1. The ratio is also reasonably consistent across different covariates. For the exacerbation to exacerbation transition,

	into exacerbation			exacerbation to exacerbation		
covariate	fixed	random	ratio	fixed	random	ratio
low dose	-0.1754	-0.2800	0.63	-0.0814	-0.0923	0.88
high dose	-0.4468	-0.5118	0.87	-0.0389	-0.0311	1.25
age	-0.0191	-0.0196	0.97	0.0061	0.0052	1.17
gender	0.2011	0.2219	0.91	-0.0273	-0.0100	2.73
time since onset	-0.0035	-0.0061	0.57	-0.0098	-0.0069	1.42
EDSS	0.0889	0.0960	0.93	0.0313	0.0347	0.90
time	-0.0051	-0.0037	1.38	0.0018	0.0015	1.20

Table 2.7: Comparison of covariate effects for first-order models

	into exacerbation		exacerbation to exacerbation	
covariate	fixed effects	random effects	fixed effects	random effects
low dose	-2.20	-1.93	-0.90	-0.89
high dose	-5.25	-3.57	-0.40	-0.24
age	-3.43	-1.93	1.01	0.80
gender	2.62	1.72	-0.31	-0.09
time since onset	-0.55	-0.55	-1.30	-0.74
EDSS	3.32	1.86	0.97	0.92
time	-4.62	-3.34	1.34	0.90

Table 2.8: Comparison of z-scores for first-order models

the values range from 0.88 to 2.73, so this is much less consistent than the other transition.

To compare the assessment of covariates in the different models we listed all the z-scores for the separate regressions in Table 2.8. Overall, the results are qualitatively quite similar for the random intercept and fixed effects model, although some of the covariates (low dose, age, gender and EDSS) which have statistically significant estimates in the fixed effects model are not significant in the random intercepts model at the 5% level (but they are still significant at the 8% level).

We performed a goodness-of-fit analysis as described in Section 2.5 for

the fixed effects model given by (2.3) and (2.4). The data were split according to the estimated no exacerbation to exacerbation transition probability at day 500. The results are shown in Figure 2.3. These plots show that the observed frequencies exhibit different behaviour in the different quartiles, and hence the departure from the estimated proportion of patients in exacerbation is also different in each quartile. Figure 2.4 shows the assessment for all the data. This plot suggests a better fit of the model than Figure 2.3, mainly because the observed frequencies show less variability over time. The estimated probabilities capture the pattern over time quite well for all the data, whereas the pattern over time is quite different in the different quartiles and hence the discrepancy between estimated occupation probabilities and the observed frequencies is considerable.

Although we have outlined in Section 2.5 how to proceed with the goodness-of-fit assessment for the random effects model, we did not carry this out, since the calculation of the estimates for the random effects is tedious.

We have noted earlier that one of the shortcomings of the Markov chain model is that the estimated effects of covariates have to be interpreted in terms of transition probabilities. However, as we have seen in the goodness-of-fit section, it is possible in principle to estimate marginal occupational probabilities (that is, the probability that a given patient is in exacerbation at time j) from the estimated transition probabilities and the response at the first time point. While this is mathematically straightforward, the fact that the starting values have a strong influence on the estimated occupation probabilities, particularly at early time points, makes this approach not very appealing. This dependence of the estimates on the initial values again results in a model where the covariates have to be interpreted conditional on the disease process. Furthermore it is not straightforward to obtain estimates for the standard errors for these estimated occupational probabilities. Muenz and Rubinstein [21] develop this approach in some detail.

The main objective for this thesis is to explore a model that incorporates the information available on the lengths of exacerbations. However, the Markov chain models presented in this chapter did not use all the information available since we collapsed the data to reduce the order of the chain. A model that takes advantage of all the information available is that of alternating re-

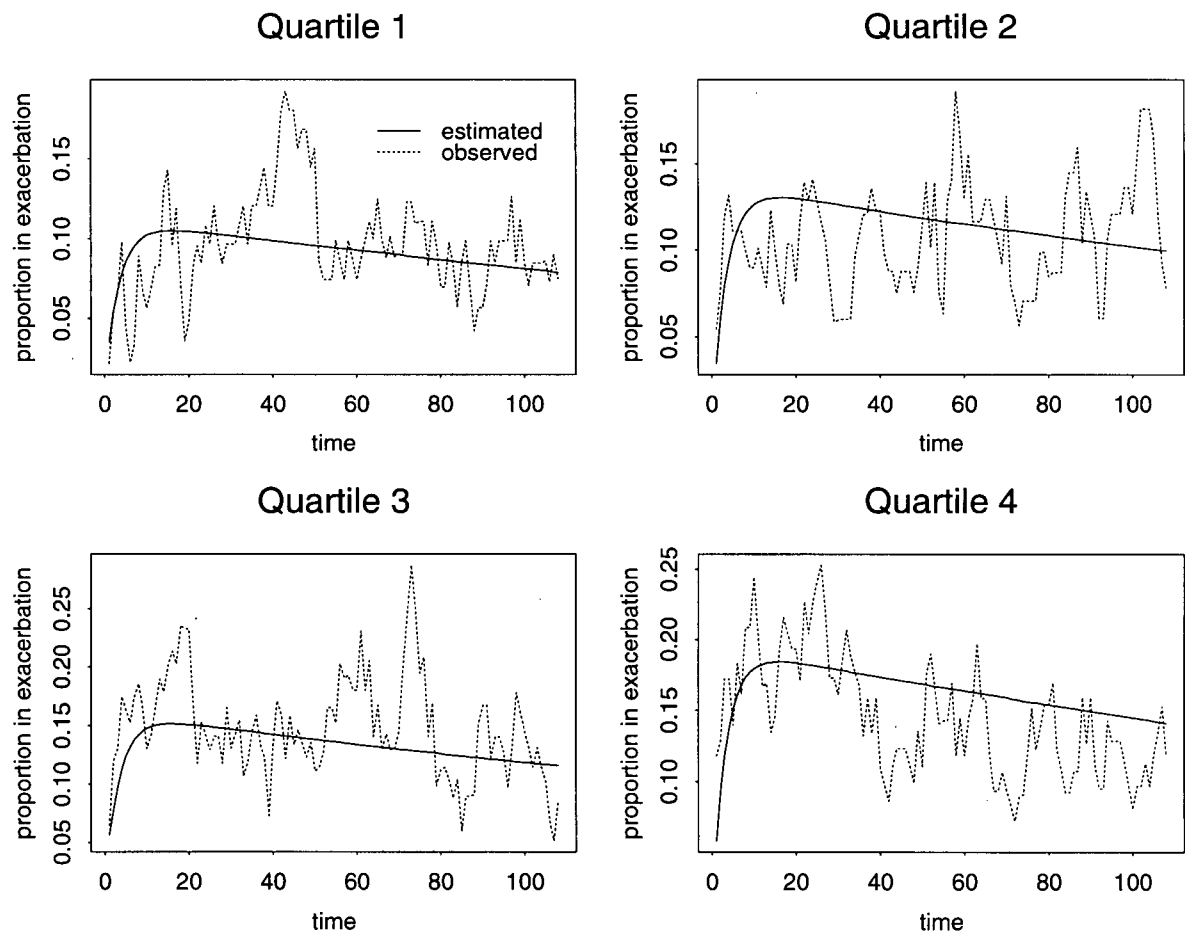


Figure 2.3: Average estimated probability and proportion of patients in exacerbation over time, by quartiles of estimated transition probability (see text for details)

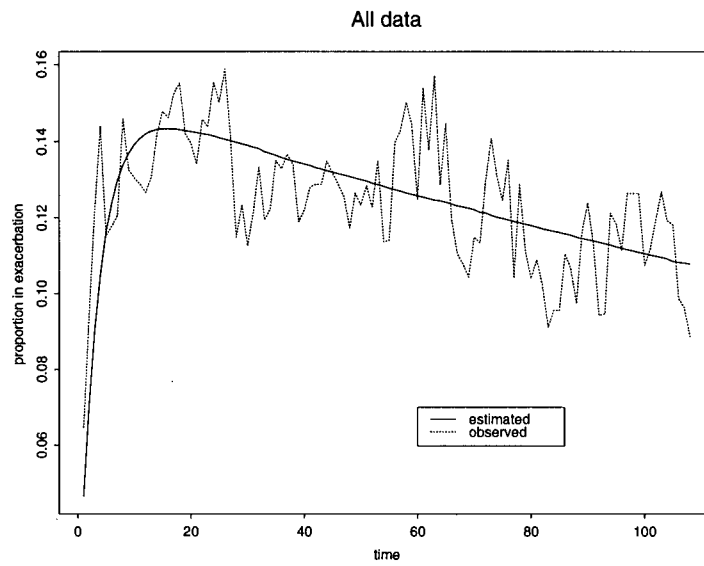


Figure 2.4: Average estimated probability and proportion of patients in exacerbation over time

new processes, which will be presented in Chapter 4.

Chapter 3

Point Processes: Basic Properties and Preliminary Analyses

This chapter has a different scope than the other chapters in this thesis. Here we do not present an approach to inference for the MS data set; rather we set the stage for the two subsequent chapters which will present analyses for the data set. In Chapter 1 we noted that alternating renewal processes and marked point processes can be expressed within the (very general) framework of point processes, so we introduce some terminology and basic properties for point processes in this chapter to avoid repetition in the subsequent chapters.

Poisson processes are an important and convenient tool for exploratory analysis for both the alternating renewal process models and the marked point process model, so we also include the exploratory analysis for the subsequent chapters here.

We do not attempt any detailed formal introduction of point processes. For this, the reader is referred to Cox and Isham [5] who cover statistical inference for point processes in some detail and Kingman [14] who gives a short introduction to Poisson processes and marked Poisson processes.

3.1 Terminology and Basic Properties

In this section, we review some general properties and terminology related to point processes. We introduce Poisson and renewal processes as important examples and show the connection to our application.

Point processes on the real line model the counts of events of a specified type in an interval. There is a close connection between the number N_t of events in $[0, t]$ and the event times T_k given by

$$\{N_t \geq n\} = \{T_n \leq t\}.$$

Alternatively one can also consider the inter-event times $T_k - T_{k-1}$, which are usually assumed to be independent.

A point process N_t can be characterized in terms of its conditional intensity function. If we denote the history of the process up to time t as $H(t) = \{N_s : s < t\}$, then the conditional intensity at time t given the history $H(t)$ is defined as

$$\lambda(t, H(t)) = \lim_{\Delta t \rightarrow 0} \frac{P(dN_t = 1 | H(t))}{\Delta t},$$

where $dN_t = 1$ if an event occurs at time t , $dN_t = 0$ otherwise. Hence we can interpret the conditional intensity function (henceforth just called conditional intensity) as the probability that an event occurs in the next instant given the prior history.

Covariates are usually included in the model by incorporating them into the conditional intensity. A convenient way to do this is to factor the intensity into a baseline intensity that depends on the history $H(t)$ and the current time t and a positive function $\phi(\mathbf{z})$ which depends upon a vector of time-independent covariates \mathbf{z} , i.e.

$$\lambda(t, H(t), \mathbf{z}) = \lambda_0(t, H(t))\phi(\mathbf{z})$$

A common choice for $\phi(\mathbf{z})$ is $\exp(\mathbf{z}'\boldsymbol{\theta})$ for a parameter vector $\boldsymbol{\theta}$.

So far we have not assumed any particular dependence of the conditional intensity on the history $H(t)$ of the process. The modeling of this dependence

defines the basic stochastic properties of the process and leads to the distinction between different kinds of processes. For Poisson processes, we incorporate only t and the covariates into the intensity so that $\lambda(t, H(t), \mathbf{z}) = \lambda(t, \mathbf{z})$; the special case where the intensity is independent of time (i.e. $\lambda(t, H(t), \mathbf{z}) = \lambda(\mathbf{z})$) is called a homogeneous Poisson process. For renewal processes, the conditional intensity for the k -th inter-event time $T_k - T_{k-1}$ depends only on the so-called *backward recurrence time* $t - T_{k-1}$ so that $\lambda(t, H(t), \mathbf{z}) = \lambda(t - T_{k-1}, \mathbf{z})$ for the k -th inter-event time. We will also use a more general class of processes called *modulated renewal processes*, in which the conditional intensity for the k -th inter-event time $T_k - T_{k-1}$ depends on the backward recurrence time and time itself so that $\lambda(t, H(t), \mathbf{z}) = \lambda(t, t - T_{k-1}, \mathbf{z})$ for the k -th inter-event time. The specification of the intensity leads to the distribution of the inter-event times $T_k - T_{k-1}$ which have density $\lambda(t, H(t), \mathbf{z}) \exp(-\int_0^t \lambda(s, H(s), \mathbf{z}) ds)$.

While the specification of the intensity leads to a distribution for the inter-event times, in applications the interest often lies in modeling the event counts at each time-point or the rate of events in a specified time interval. Poisson processes are particularly attractive for this situation, since they provide a simple connection between the intensity and the mean counts given by

$$E(N_t | \mathbf{z}) = \int_0^t \lambda(s, \mathbf{z}) ds.$$

In the special case of an homogeneous Poisson process this leads to

$$E(N_t | \mathbf{z}) = \lambda(\mathbf{z})t.$$

Hence the incorporation of covariates into the intensity leads to an easy interpretation of covariates with respect to the mean counts and the rate of occurrence of events (given by $\frac{d}{dt}E(N_t | \mathbf{z})$) is also the intensity. Homogeneous Poisson processes also provide a simple connection between the lengths of intervals between events and the intensity. More precisely, if the process has the constant intensity $\lambda(\mathbf{z})$, then the inter-event times have an exponential distribution with mean $\lambda(\mathbf{z})^{-1}$ (and hence variance $\lambda(\mathbf{z})^{-2}$).

Our objective is to build models for a two-state process, but the point processes considered so far do not distinguish between different events and

hence are not general enough. To incorporate this second state into the analysis, we use two models: alternating renewal processes and marked point processes. In the remainder of this section we will give an introduction to both models and show the connection to the application to the MS data set.

Alternating renewal processes assume that an individual process alternates between two states. An event is the change of state and the intensity for the next event depends on the time that has elapsed since the process entered the current state. This means that to incorporate the second state we simply need to model two transition intensities. A bit more formally, we denote the event times for the state 1 to state 2 transitions by T_k and the conditional intensity associated with this transition by λ_1 . Similarly we denote the event times for the state 2 to state 1 transition by S_k and the associated conditional intensity by λ_2 . In this way we can represent the process as a sequence of event times $(T_1, S_1, T_2, S_2, \dots)$ or $(S_1, T_2, S_2, T_3, \dots)$, depending on whether the process starts in state 1 or state 2 (we omit T_1 in the latter case for notational convenience). We then assume that the inter-event time $S_k - T_k$ (during this time the process is in state 2) has intensity $\lambda_2(t, t - T_k, \mathbf{z})$ and the inter-event time $T_k - S_{k-1}$ where the process is in state 1 has intensity $\lambda_1(t, t - S_{k-1}, \mathbf{z})$. In the application to the MS data, state 1 is “not in exacerbation”, state 2 is “in exacerbation” and an event is hence the beginning or the end of an exacerbation.

Marked point processes take a different approach. For this model we assume that we have a sequence of events (as for any point process), but each event has a so-called mark associated with it. This means we can describe the process by a sequence of event times T_k combined with a sequence of marks X_k , with the k -th mark X_k associated with the k -th event time T_k . Hence we have an underlying point process N_t given by the sequence (T_1, T_2, \dots) and a process of marks $M_t = \sum_{\{i: 0 < T_i \leq t\}} X_i$. They are related by the assumption that, given $N_t = n$, M_t is a sum of n independent random variables. For a more detailed discussion and more general examples, see [5], p. 131. In the application to the MS data set an event is the beginning of an exacerbation and the mark associated with this event is the length of this exacerbation.

3.2 Initial Data Analysis

In this section we first show that we can use the same initial data analysis for the alternating renewal process model and the marked point process model. We then proceed with the initial data analysis for the MS data set, which is relevant to both of the next two chapters.

We have seen in the previous section that we need to model :

- the conditional intensity for transition into and out of exacerbation for alternating renewal processes
- the conditional intensity for transition into exacerbation and length of exacerbations for marked point processes.

We have also seen that for a point process with constant intensity λ (which is then also the rate) the inter-event times have mean $1/\lambda$. Hence for purposes of data exploration we can use alternatively the mean inter-event times or the rate of events (i.e. the number of events in a specified time period). Hence we use the yearly rate of exacerbations and the mean length of time spent in exacerbation for both approaches.

We are interested how the baseline covariates are related to these two outcomes (rate of exacerbation and mean exacerbation length). To explore this relationship, we summarize the data on each individual and then plot the outcomes versus the covariates. We summarize the data for each patient for state 1 as yearly exacerbation rates (number of exacerbations divided by the time on study, rescaled by 365 to obtain the rate per year) and the data for the exacerbations as the average lengths of exacerbations (if the patient did not have an exacerbation, this summary is missing).

Figure 3.1 shows the individual rates versus the baseline covariates. We see that a number of patients have fairly high yearly rates (up to 12, but this patient, number 418, dropped out early), and there is a substantial amount of variation among individual rates. We notice that the rates in the placebo group seem to be higher than in the low dose and high dose groups, but there is no apparent difference between the low dose and high dose group. The time since onset seems to be correlated with the rates decreasing as the time since

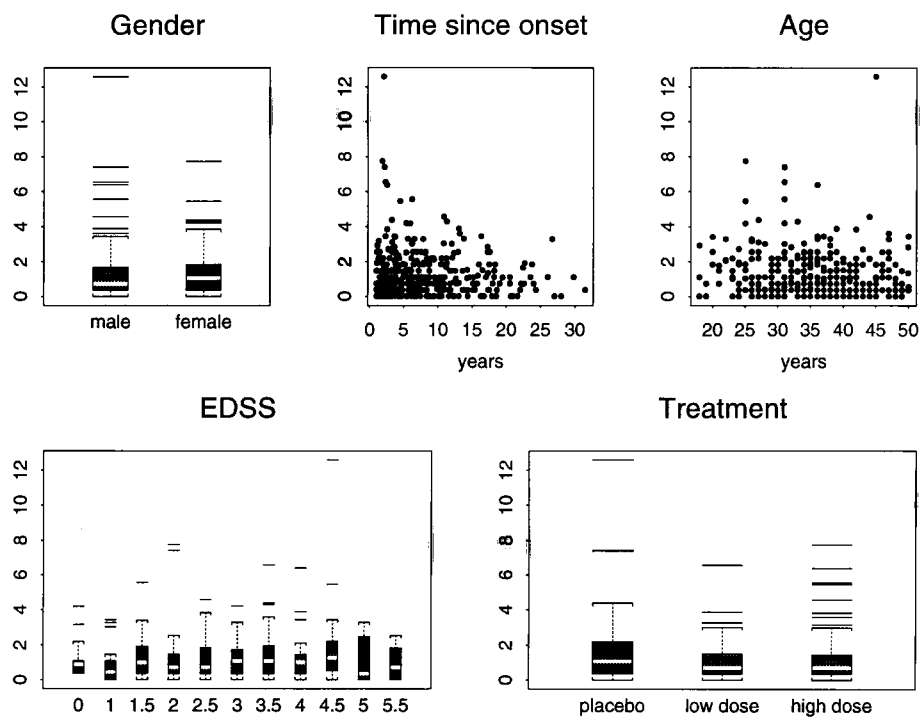


Figure 3.1: Rate of exacerbations versus covariates

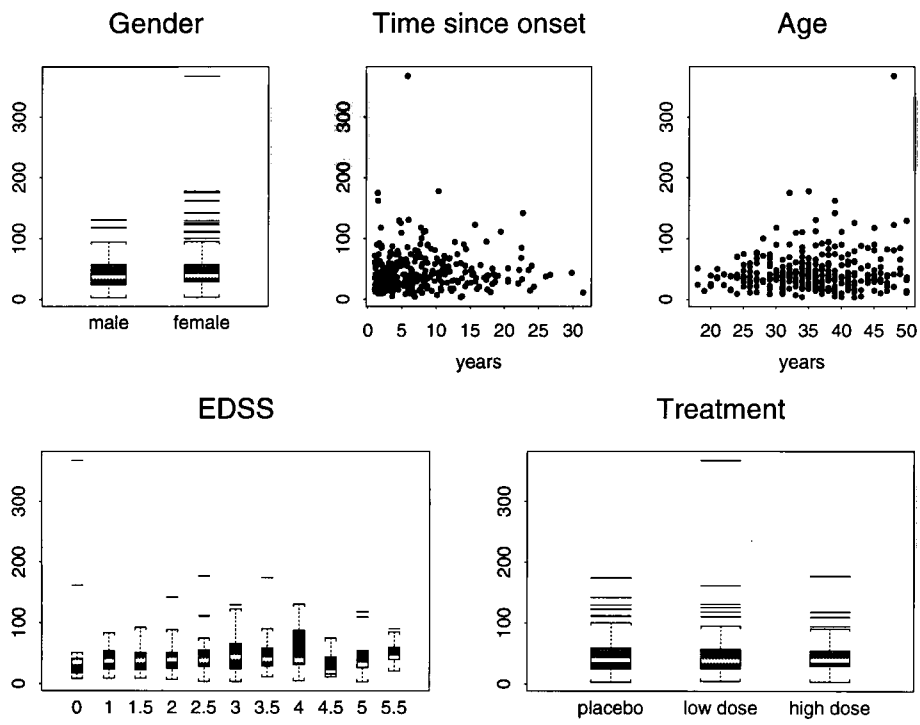


Figure 3.2: Average lengths of exacerbations versus covariates

onset increases. There is an indication that female patients have higher exacerbation rates. The plots for EDSS and age do not show any particular pattern.

Figure 3.2 shows the individual average exacerbation lengths versus the baseline covariates. There is a striking variability in the average lengths of exacerbations across individuals with one extreme observation (outlier) of more than 300 days (this patient, number 502, had one exacerbation 368 days long). There is not much indication of a treatment effect, and there is no apparent relationship between EDSS and the length of exacerbations. It is worth noting that the patient with the extreme average length has an EDSS of zero, which could mean that this point has high leverage. The median average exacerbation length for male and female patients is almost equal, but there are more extreme observations for the female patients. There seems to be a tendency for shorter exacerbations among patients who had MS for a longer time, and a tendency for longer exacerbations among older patients. Note also that the outlier amplifies the apparent relationship for both the time since onset and age.

We are also interested in changes in these outcomes (rates of exacerbations and average exacerbation lengths) over time, and in particular how the treatment effect changes over time. To explore this, we split the time axis into 5 intervals of 200 days (0-200, 201-400, 401-600, 601-800 and 801-1000) and calculated the rate of exacerbations and the average lengths of exacerbations for exacerbations starting in this time period for each treatment group and time period.

Figure 3.3 shows the rates by treatment group over time. This plot shows a marked difference between the placebo and high dose groups so that there is apparently a treatment effect with the high dose treatment. There is some hint that the low dose treatment is effective as well, but the difference from placebo group is much less apparent. There is also a clear decrease over time of the rates in all treatment arms. There is some indication of a treatment by time interaction for the rates, since the lines are not parallel. However, this effect does not seem to be very strong.

We are also interested in the average length of exacerbations over time.

	Time Period (days on study)				
treatment group	0-200	201-400	401-600	601-800	801-1000
placebo	108	75	71	59	43
low dose	83	76	63	52	44
high dose	65	54	46	44	37

Table 3.1: Number of exacerbations by treatment and time-period

However, we have seen there are some extreme observations for the average length of exacerbations. To see if these influence the patterns over time unduly, we also included a plot of the median average exacerbation length by time-period. Figure 3.4 shows the mean and median average exacerbation length by treatment arm over time. This plot shows that there is apparently no consistent pattern over time for the different treatment arms, neither for the median nor for the mean. Comparing these two plots we can see that the extreme observations indeed have quite a strong influence on the average time in exacerbation.

Part of the reason that this plot is so erratic could also be that the number of patients who experience an exacerbation is decreasing over time as noted earlier, so that the total number of exacerbations is decreasing over time, which can be seen in Table 3.1.

The next two chapters implement the alternating renewal process and marked point process models for the MS data.

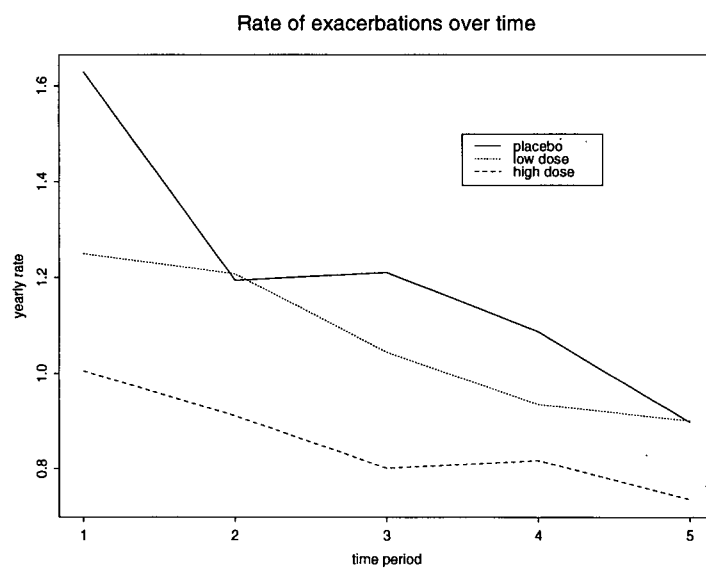
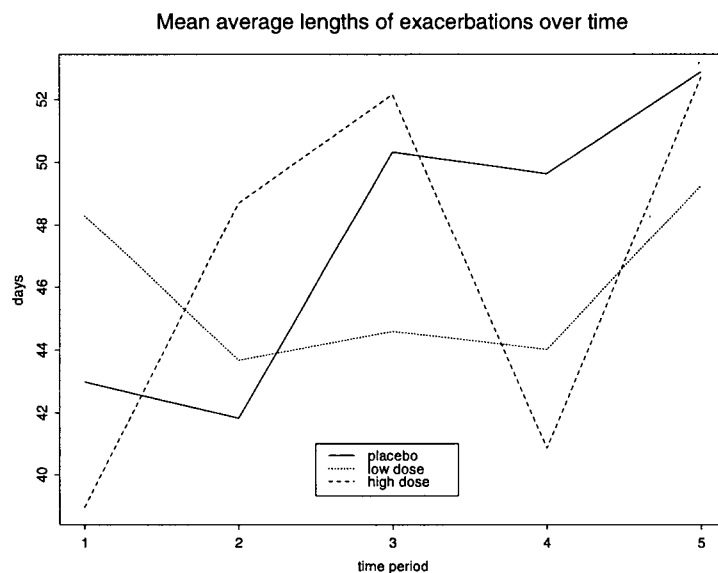
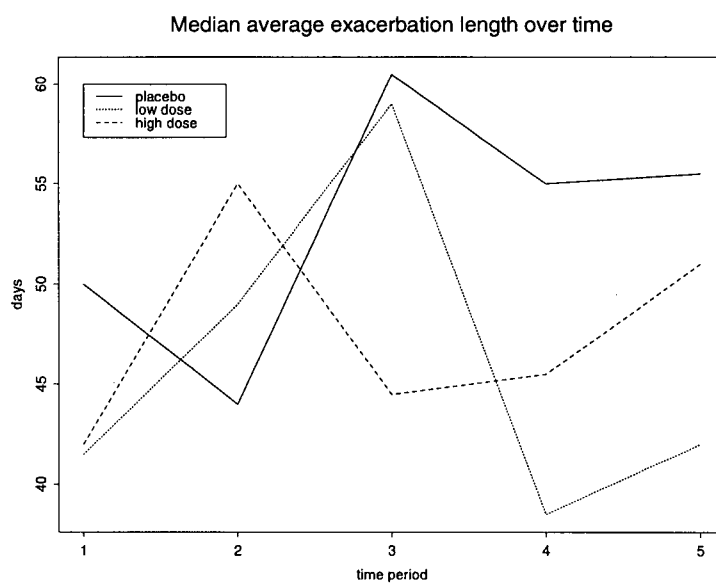


Figure 3.3: Rate of exacerbation over time in 200 day time periods by treatment



(a) Mean average lengths of exacerbations



(b) Median average lengths of exacerbations

Figure 3.4: Mean and median average lengths of exacerbations over time in 200 day time periods by treatment

Chapter 4

An Alternating Renewal Process Model

4.1 Introduction

The previous chapter introduced basic properties of point processes and set the stage for this chapter, which deals with the implementation of the alternating renewal process model. We have seen in Chapter 2 that the Markov chain model has the disadvantage that it is not easy to incorporate the lengths of exacerbations directly. In Chapter 3 we have seen that renewal processes incorporate the backward recurrence time, hence this approach allows us to deal directly with the length of exacerbations.

For the alternating renewal process approach, we model each individual disease process as an alternating renewal process. The patient alternates between state 1 (not in exacerbation) and state 2 (in exacerbation) and we record when the transition between these states occurred and how long the patient stayed in each state. Hence we can use the exacerbation data from the trial as it was recorded.

As outlined in Chapter 3, we model the intensity for the waiting time to the next event (events here are a change of state) given the covariates (including time since randomization) and the time since the last event (the backward recurrence time). One expects the backward recurrence time to be an important predictor, particularly for the time in exacerbation, since these

exacerbations usually last at least a few days, but not very long. For the times where patients are not in exacerbation, one expects this effect to be less apparent due to the large individual differences in the length of exacerbations.

The time since randomization is also of interest. The data set at hand is for relapsing-remitting MS patients. These patients eventually enter the secondary progressive stage where the disease process is very different, so one wants to allow for changes over time in the transition probabilities. Cox [4] named this kind of renewal process with multiple time scales a modulated renewal process. Other papers that deal with statistical inference for these processes are Lawless and Thiagaraja [19] and Oakes and Cui [25].

One should note that for repeated events the backward recurrence time plays the role of “time” in survival analysis where there is at most one event per individual. We also incorporate time since randomization as a covariate into our model, but in contrast to survival analysis this does not mainly describe the hazard function for the waiting times, but rather the changes for this hazard over time (i.e. how an early exacerbation differs from a late one).

Initially, we develop a fixed effects model. To accommodate heterogeneity between the individuals and to allow for correlation between the multiple waiting times on one individual (both for waiting times in one state as well as for the correlation of waiting times in state 1 with the waiting times in state 2), we introduce bivariate random intercepts into the fixed effects model. One interpretation of random effects is that they account for unobserved covariates which cause both heterogeneity between different individuals and correlation within an individual. Our approach follows closely the paper by Ng and Cook [23]. Another paper that uses bivariate random effects is Xue and Brookmeyer [34]. Aalen and Husebye [1] discuss random effects models for renewal processes and Lawless [16] discusses random effects for Poisson process models.

We use the subscripts $i = 1, \dots, m$ for the individuals, $j = 1, 2$ for the state (1 = not in exacerbation, 2 = in exacerbation). Let $t_{i1} < \dots < t_{in_i} < t_{in_i+1}$ be the observed transition times and the censoring (dropout) time for individual i . We denote the history of subject i up to time t by $H_i(t)$, including the covariate vectors $\mathbf{x}_{i1}, \mathbf{x}_{i2}$ for states 1 and 2 respectively; that is, if t_{ik} is

the last event time prior to t , we have

$$H_i(t) = \{\mathbf{x}_{i1}, \mathbf{x}_{i2}, t_{i1}, \dots, t_{ik}\}$$

4.2 Fixed Effects

In this section, we introduce a fixed effects model. It assumes the independence of inter-event times and hence may be too simple a model for a longitudinal data set, but it is the building block for the random effects model considered in the next section.

4.2.1 Parameterization of the Intensities

As mentioned in the previous chapter, we use the conditional intensities into exacerbation and out of exacerbation to model the alternating renewal process. Let λ_{ij} denote the intensity for individual i for the transition out of state j , with $\boldsymbol{\theta}_j = (\boldsymbol{\beta}_j, \boldsymbol{\xi}_j)$ the vector of parameters for state j . Here $\boldsymbol{\beta}_j$ is the set of regression parameters related to the time-independent covariates and $\boldsymbol{\xi}_j$ is the set of parameters associated with the baseline intensity λ_{0j} . The intensity is modeled as

$$\lambda_{ij}(t, \boldsymbol{\theta}_j | H_i(t)) = \exp(\mathbf{x}'_{ij} \boldsymbol{\beta}_j) \lambda_{0j}(t, \boldsymbol{\xi}_j | H_i(t)). \quad (4.1)$$

We describe the parametrization for the baseline intensity in Section 4.4.

4.2.2 The Likelihood

Let $t_{i1} < \dots < t_{in_i} < t_{in_i+1}$ be the observed transition times and the censoring time for individual i (with abuse of notation, we denote the corresponding random variables by t_{ik} as well). The inter-event times $t_{ik} - t_{i,k-1}$, $k = 1, \dots, n_i + 1$, are assumed to be independent and t_{ik} , given $t_{i,k-1}$ has density

$$f(t | H_i(t)) = \lambda_{ij}(t, \boldsymbol{\theta}_j | H_i(t)) \exp \left\{ - \int_{t_{i,k-1}}^t \lambda_{ij}(s, \boldsymbol{\theta}_j | H_i(s)) ds \right\}$$

(if $t_{ik} - t_{i,k-1}$ corresponds to a sojourn time in state j) and survival function

$$S(t | H_i(t)) = \exp \left\{ - \int_{t_{i,k-1}}^t \lambda_{ij}(s, \boldsymbol{\theta}_j | H_i(s)) ds \right\}.$$

Then the likelihood for individual i is (denoting the vector of parameters for both states by $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$)

$$L_i(\boldsymbol{\theta}) = \prod_{j=1}^2 \left\{ \prod_{k \in D_{ij}} f(t_{ik} | H_i(t_{ik})) S(t_{in_i+1} | H_i(t_{in_i+1})) 1_{\{n_i+1 \in D_{ijc}\}} \right\}. \quad (4.2)$$

Here D_{ij} is the set of indices $k \leq n_i$ where subject i is in state j for $t_{i,k-1} \leq t \leq t_{ik}$ and D_{ijc} is $D_{ij} \cup \{n_i + 1\}$ if the censoring occurred in state j and otherwise just D_{ij} . Plugging in the formulas for λ_{ij} , we get

$$L_i(\boldsymbol{\theta}) = \prod_{j=1}^2 \prod_{k \in D_{ij}} \exp(\mathbf{x}'_{ij} \boldsymbol{\beta}_j) \lambda_{0j}(t) \exp \left(- \sum_{k \in D_{ijc}} e^{\mathbf{x}'_{ij} \boldsymbol{\beta}_j} \Lambda_{0j}(t_{i,k-1}, t_{ik}) \right)$$

where

$$\Lambda_{0j}(a, b) = \int_a^b \lambda_{0j}(s) ds$$

is the integrated baseline intensity. By the independence of data from different individuals, the likelihood for m individuals $L(\boldsymbol{\theta})$ is then

$$L(\boldsymbol{\theta}) = \prod_{i=1}^m L_i(\boldsymbol{\theta}).$$

4.3 Random Effects

The fixed effects model assumes that the sojourn times on an individual are independent, so does not allow for correlation of the data on an individual. To accommodate this, we expand the fixed effects model introduced above by introducing random intercepts (u_{i1}, u_{i2}) in the model specification (4.1) for the intensities, i.e.

$$\lambda_{ij}(t, \boldsymbol{\theta}_j | H_i(t), u_{ij}) = \exp(\mathbf{x}'_{ij} \boldsymbol{\beta}_j + u_{ij}) \lambda_{0j}(t, \boldsymbol{\xi}_j | H_i(t)). \quad (4.3)$$

By assuming that the individual intercepts for each state are a random sample, we introduce a positive correlation between the sojourn times in state 1 and between the sojourn times in state 2 on each individual. In addition, we assume that the random effects are correlated; that is, that (u_{i1}, u_{i2}) are a

random sample from a bivariate distribution. This allows a correlation for the sojourn times in state 1 with the sojourn times in state 2. Unfortunately, the correlation of the transition intensities does not easily translate into a correlation of the sojourn times (see Lindeboom and Van Den Berg, [20]) so that the correlation for the intensities is not easily interpretable. But the sign of the correlation is the same for the transition intensities as for the sojourn times, so the sojourn times can be negatively correlated and the sign of the estimated correlation for the intensities can be used for a qualitative assessment of the correlation of the sojourn times.

4.3.1 The Distribution of the Random Effects

We assume that the intercepts (u_{i1}, u_{i2}) follow a bivariate normal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and unknown covariance matrix $\Sigma = (\sigma_{kl})$. The parametrization (4.3) then implies that, at any fixed time t , the intensities $(\lambda_{i1}(t, \boldsymbol{\theta}_j | H_i(t), u_{ij}), \lambda_{i2}(t, \boldsymbol{\theta}_j | H_i(t), u_{ij}))$ follow a bivariate log-normal distribution. This choice for the distribution of the random effects differs from the usual choice for random effects for repeated failure times; we discuss this at the end of this chapter. We now introduce some general properties of the bivariate log-normal distribution.

Let $\mathbf{U} = (U_1, U_2)$ be bivariate normal with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and covariance matrix $\Sigma = (\sigma_{kl})$. Then $(X_1, X_2) = (\exp(U_1), \exp(U_2))$ is bivariate log-normal and we have

$$\begin{aligned} E(X_1^{r_1} X_2^{r_2}) &= E(\exp(\mathbf{r}^T \mathbf{U})) \\ &= \exp(\mathbf{r}^T \boldsymbol{\mu} + \frac{1}{2} \mathbf{r}^T \Sigma \mathbf{r}) \end{aligned}$$

where the last equation follows from the expression for the moment generating function of a bivariate normal random variable. In our application, we require the mean of \mathbf{U} to be $\boldsymbol{\mu} = (0, 0)$. In this case, we have

$$E(X_j) = \exp\left(\frac{1}{2} \sigma_{jj}\right). \quad (4.4)$$

In addition we have

$$\text{cov}(X_1, X_2) = \exp\left(\frac{1}{2}(\sigma_{11} + \sigma_{22})\right) (\exp(\sigma_{12}) - 1)$$

and

$$\text{cov}(X_j, X_j) = \exp(\sigma_{jj})(\exp(\sigma_{jj}) - 1)$$

for $j = 1, 2$. Thus

$$\text{corr}(X_1, X_2) = \frac{\exp(\sigma_{12}) - 1}{\sqrt{(\exp(\sigma_{11}) - 1)(\exp(\sigma_{22}) - 1)}}. \quad (4.5)$$

We mentioned above that the transition intensities have a bivariate log-normal distribution. Hence (4.5) shows that a negative correlation of the intercepts leads to a negative correlation of the transition intensities.

4.3.2 The Likelihood

The derivation of the likelihood for the fixed effects model depends on the assumption that the inter-event times are independent. In the random effects model, we assume that the inter-event times $t_{ik} - t_{i,k-1}$, $k = 1, \dots, n_i + 1$, are independent conditional on \mathbf{u}_i . Furthermore, t_{ik} given $t_{i,k-1}$ has density

$$f(t|H_i(t), u_{ij}) = \lambda_{ij}(t, \boldsymbol{\theta}_j | H_i(t), u_{ij}) \exp \left\{ - \int_{t_{i,k-1}}^t \lambda_{ij}(s, \boldsymbol{\theta}_j | H_i(s), u_{ij}) ds \right\}$$

so that, in analogy to (4.2) the conditional likelihood for individual i is

$$L_i(\boldsymbol{\theta} | \mathbf{u}_i) = \prod_{j=1}^2 \prod_{k \in D_{ij}} \exp(\mathbf{x}'_{ij} \boldsymbol{\beta}_j + u_{ij}) \lambda_{0j}(t) \exp \left(- \sum_{k \in D_{ijc}} e^{(\mathbf{x}'_{ij} \boldsymbol{\beta}_j + u_{ij})} \Lambda_{0j}(t_{ik-1}, t_{ik}) \right). \quad (4.6)$$

The marginal likelihood is obtained by integrating with respect to the density g of the random effects (u_{i1}, u_{i2}) . In our application, g is the bivariate normal density with mean $(0,0)$ and unknown covariance matrix Σ . Since the data and the random effects for different individuals are assumed to be independent, the marginal likelihood is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^m \int L_i(\boldsymbol{\theta} | \mathbf{u}) g(\mathbf{u}) d\mathbf{u}. \quad (4.7)$$

4.4 Parameterization of the Baseline Intensity

Like Ng and Cook [23], we also consider two time scales as covariates in the model: the time since randomization and the backward recurrence time. We refer to the first as the “Markov” time, since it represents the history on this time scale for this individual. We refer to the backward recurrence time as the “Renewal” time scale. This is in analogy to renewal processes, where the conditional transition probabilities given the complete history depend only on the time since the last event. In a disease process like relapsing-remitting MS, these time scales have different interpretations. The renewal scale represents the dynamics of the waiting times in one state: Given the time spent in this state, how likely is it to switch to the other state in the next instant? The Markov scale on the other hand describes the disease progression: How does the transition probability after three years differ from that shortly after randomization?

These time-scales are included in the baseline intensity, which we parametrize as the product of the Markov component T_j and the renewal component R_j

$$\lambda_{0j}(t, \boldsymbol{\xi}_j | H_i(t)) = R_j(b_i(t); \boldsymbol{\gamma}_j) T_j(t; \boldsymbol{\delta}_j)$$

where $b_i(t)$ denotes the backward recurrence time and $\boldsymbol{\xi}_j = (\boldsymbol{\gamma}_j, \boldsymbol{\delta}_j)$.

We consider three different parametrizations for the two time-scales, separately for each state:

- linear:

$$\begin{aligned} R_j(s) &= \exp(\boldsymbol{\gamma}_j s) \\ T_j(s) &= \exp(\boldsymbol{\delta}_j s) \end{aligned}$$

- piecewise constant:

$$\begin{aligned} R_j(s) &= \sum_{l=1}^{r_j} \exp(\boldsymbol{\gamma}_{jl}) 1_{\{a_{jl} \leq s < a_{j,l+1}\}} \\ T_j(s) &= \sum_{l=1}^{m_j} \exp(\boldsymbol{\delta}_{jl}) 1_{\{b_{jl} \leq s < b_{j,l+1}\}} \end{aligned}$$

where a_{jl} and b_{jl} are known constants specifying the intervals on which the functions are constant.

- with cubic B-splines $sp^{R,j}(s; \gamma_j), sp^{M,j}(s, \delta_j)$:

$$\begin{aligned} R_j(s) &= \exp(sp^{R,j}(s; \gamma_j)) \\ T_j(s) &= \exp(sp^{M,j}(s, \delta_j)). \end{aligned}$$

We consider different parametrizations for the time-scales for two reasons: to check the sensitivity of the inference for the baseline covariates on the parametrization of the time-scales and to allow a flexible dependence of the transition intensities on time to examine the patterns over time. We don't expect the dependence on the backward recurrence time to be linear; we mainly considered the linear parametrization for this time-scale to see if this overly simple model still leads to correct inference for the baseline covariates. For the Markov time-scale, the linear parametrization might not be unreasonable

Two of these parametrizations allow a flexible pattern over time: the piecewise constant and cubic B-splines parametrizations. The piecewise constant parametrization is a flexible approach that allows us to investigate effects over time without assuming a complicated functional form for these effects. It is easy to fit with relatively few parameters, but assumes a step function for the estimated dependence on time which may not be the functional form one expects for these effects. Splines, on the other hand, model the dependence over time with flexible, smooth functions. This parametrization has the advantage that it is a more plausible model for a disease process as we expect the dependence of the transition probabilities to be smooth over time. The parametrization is more complicated to implement however, and we encountered numerical difficulties with this approach.

In the remainder of this section we outline our implementation of the spline parametrization; the implementation of the other parametrizations is straightforward. To simplify notation, we drop all subscripts for the state or time-scale and use generic parameters and functions instead; so for example the l -th basis function $b_l(s)$ (introduced below) represents one of the four basis functions $b_l^{R,1}(s), b_l^{M,1}(s), b_l^{R,2}(s)$ or $b_l^{M,2}(s)$.

A B-spline $sp(s, \gamma)$ on the interval $[a, b]$ is a linear combination of basis functions $b_l(s)$, say

$$sp(s, \gamma) = \sum_{l=1}^p \gamma_l b_l(s).$$

The number and the shape of the basis functions $b_l(s)$ depend on the number and values of the knots $s_l \in [a, b]$ which have to be chosen first. These basis functions are defined such that $sp(s, \gamma)$ has continuous second derivatives on $[a, b]$, and is a polynomial on each of the intervals $[s_l, s_{l+1}]$. The most important feature of splines for our implementation is that once the number and location of the knots is fixed, the spline is completely determined by the vector of parameters γ . Possible choices for the knots s_k are, for example, quantiles from the sample of observed Markov or renewal times; this ensures that each of the intervals $[s_k, s_{k+1}]$ contains the same number of data points.

In the application to the MS data set, for each state (1,2) and each time-scale (Markov, renewal) we first choose a set of knots. The splines are then completely determined by the parameter vectors $\gamma_1, \delta_1, \gamma_2$ and δ_2 . We then include these four parameter vectors in the procedure to maximize the likelihood.

Choosing the knots prior to the maximization of the likelihood is in contrast to the usual statistical use of splines, where the choice of the knots is also optimized in some way. We felt, however, that for our purposes this simpler approach would suffice. Our use of splines is intended to allow for a flexible functional form, but we do not necessarily want to obtain the “best” approximation to the relationship between time and transition intensities. The main advantage of fixing the knots in advance is that the implementation is relatively straightforward and does not require many changes from the implementation for other parametrizations for the time-scales.

4.5 Numerical Methods

4.5.1 Random Effects

As a part of the estimation, we have to derive estimates for the entries of the covariance matrix of the random effects:

$$\text{Var}(u_{ij}) = \Sigma = \begin{pmatrix} \sigma_{11} & \sqrt{\sigma_{11}\sigma_{22}\rho} \\ \sqrt{\sigma_{11}\sigma_{22}\rho} & \sigma_{22} \end{pmatrix}.$$

The estimation must take the constraints on the parameters into account: the variances have to be positive and the correlation between -1 and 1 . To ensure these constraints are satisfied, we use the Cholesky decomposition to re-parametrize $\Sigma = \Omega\Omega^T$ where

$$\Omega = \begin{pmatrix} \omega_1 & 0 \\ \omega_3 & \omega_2 \end{pmatrix}.$$

Then $\sigma_{11} = \omega_1^2$, $\sigma_{22} = \omega_2^2 + \omega_3^2$ and $\rho = \omega_3/\sqrt{\omega_2^2 + \omega_3^2}$ and there are no restrictions on the vector $\omega = (\omega_1, \omega_2, \omega_3)$. We re-parametrize the random effects as $\mathbf{z}_i = \Omega^{-1}\mathbf{u}_i$ where $\mathbf{z}_i \sim N(0, I_2)$ and I_2 is the 2×2 identity matrix. The conditional intensity then becomes

$$\lambda_{ij}(t, \boldsymbol{\theta}_j | H_i(t), z_{ij}) = \exp(\mathbf{x}'_{ij}\boldsymbol{\beta}_j + \Omega_j \mathbf{z}_i) \lambda_{0j}(t, \boldsymbol{\xi}_j | H_i(t))$$

where Ω_j is the j th row of Ω . The marginal likelihood for subject i thus becomes

$$\begin{aligned} L_i(\boldsymbol{\theta}, \omega) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} L_i(\boldsymbol{\theta}, \omega | \mathbf{z}_i) d\Phi(z_{i1}) d\Phi(z_{i2}) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} L_i(\boldsymbol{\theta}, \omega | \mathbf{z}_i) e^{-\frac{1}{2}z_{i1}^2} e^{-\frac{1}{2}z_{i2}^2} dz_{i1} dz_{i2} \\ &= \frac{1}{2\pi} \left(\prod_{j=1}^2 \prod_{k \in D_{ij}} \exp(\mathbf{x}'_{ij}\boldsymbol{\beta}_j) \lambda_{0j}(t) \right) \\ &\quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left\{ n_{i1}\omega_1 z_{i1} + n_{i2}(\omega_2 z_{i2} + \omega_3 z_{i1}) \right\} - \\ &\quad \sum_{j=1}^2 \sum_{k \in D_{ijc}} e^{(\mathbf{x}'_{ij}\boldsymbol{\beta}_j + \Omega_j z_{ij})} \Lambda_{0j}(t_{ik-1}, t_{ik}) \Big\} e^{-\frac{1}{2}z_{i1}^2} e^{-\frac{1}{2}z_{i2}^2} dz_{i1} dz_{i2} \end{aligned}$$

where n_{ij} is the number of uncensored transitions out of state j .

To maximize the log-likelihood for m individuals, given by (see (4.7))

$$\log L(\boldsymbol{\theta}, \omega) = \sum_{i=1}^m \log L_i(\boldsymbol{\theta}, \omega),$$

we use a quasi-Newton method. That is, the parameter vector in the $(k+1)$ st iteration is given by

$$\begin{pmatrix} \boldsymbol{\theta}^{(k+1)} \\ \omega^{(k+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\theta}^{(k)} \\ \omega^{(k)} \end{pmatrix} - \left(\Delta^2 \log L(\boldsymbol{\theta}^{(k)}, \omega^{(k)}) \right)^{-1} \Delta \log L(\boldsymbol{\theta}^{(k)}, \omega^{(k)})$$

where $\Delta \log L(\boldsymbol{\theta}, \omega)$ is the finite difference approximation to the first derivative of the log-likelihood (i.e. the score vector) and $\Delta^2 \log L(\boldsymbol{\theta}, \omega)$ is the corresponding approximation to the second derivative of the log-likelihood. Here the vector Δf for a function $f : R^p \rightarrow R$ and a suitably chosen small constant h is given by

$$(\Delta f(a_1, \dots, a_p))_l = \frac{f(a_1, \dots, a_l + h, \dots, a_p) - f(a_1, \dots, a_l - h, \dots, a_p)}{2h},$$

and the (l, m) th element of the matrix $\Delta^2 f$ is given by

$$\begin{aligned} & (\Delta^2 f(a_1, \dots, a_p))_{l,m} \\ &= \{ [f(a_1, \dots, a_l + h, \dots, a_m + h, \dots, a_p) - f(a_1, \dots, a_l + h, \dots, a_m - h, \dots, a_p)] \\ & - [f(a_1, \dots, a_l - h, \dots, a_m + h, \dots, a_p) - f(a_1, \dots, a_l - h, \dots, a_m - h, \dots, a_p)] \} / 4h^2. \end{aligned}$$

4.5.2 Splines

The evaluation of the splines for implementation in C caused some complications. Initially we thought that we could use the functions supplied in S-plus to evaluate the basis functions. While this is possible in principle, Venables and Ripley [33] suggest the use of a look-up table instead to achieve higher speed.

To implement this, we create vectors of values for the different basis functions in S-plus and import them into C (the use of a fixed set of knots makes this feasible). This gives us the values at a set of points; we use linear

interpolation to approximate these basis functions at intermediate points. By increasing the number of points that we import into C, we can improve this approximation (at the expense of speed, since for each evaluation of the spline we have to locate the appropriate values in the look-up tables).

4.6 Comments on the Numerical Method

Our implementation of the maximization routine to obtain the MLE differs from that suggested by Ng and Cook [23]. They use Gauss-Hermite integration to obtain weights w_l, w_m and points a_l, a_m for $l, m = 1 \dots N$ at which to evaluate the integrand (often called abscissas in the numerical analysis literature), to approximate an integral of the form

$$\log \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(z_1, z_2) e^{-z_1^2} e^{-z_2^2} dz_1 dz_2 \right)$$

with a sum

$$\log \left(\sum_{l=1}^N \sum_{m=1}^N g(a_l, a_m) w_l w_m \right) \quad (4.8)$$

where

$$g(z_1, z_2) = L_i(\boldsymbol{\theta}, \omega | z_1, z_2)$$

for individual i . These integrals appear in the marginal log-likelihood function. To implement the Newton-Raphson algorithm, they simply approximate the log-likelihood using (4.8) and then analytically differentiate the approximation.

We chose to use a different approach for two reasons. First, the second derivatives of the log-likelihood are quite complicated analytical expressions, so we prefer a quasi-Newton routine that is less prone to programming errors and also much easier to change if one wants to implement a different model (e.g. with splines). Second, we found that the Ng and Cook approach was quite unstable numerically, for reasons that we now explain.

To simplify the discussion, consider the case of a one-dimensional function $g(z)$ and the integral

$$\int_{-\infty}^{\infty} g(z) e^{-z^2} dz.$$

Gauss-Hermite integration selects “optimal” weights w_m and abscissas a_m for $m = 1, \dots, N$ to obtain the approximation

$$\int_{-\infty}^{\infty} g(z) e^{-z^2} dz \approx \sum_{m=1}^N g(a_m) w_m.$$

These weights and abscissas are optimal in the sense that this approximation is exact for a polynomial of degree $\leq 2N - 1$, and works well if g is well-approximated by a polynomial (for further discussion, see Press et al. [26]; expressions for the error are given in Stoer and Bulirsch [31]).

But the function we have to integrate is very different from a polynomial. For independent random effects (i.e. $\omega_3 = 0$), the part of the likelihood we have to integrate factors into

$$\exp(n_1 \omega_1 z_1 - e^{\omega_1 z_1} e^{x_1 \beta_1} \Lambda_1) \exp(n_2 \omega_2 z_2 - e^{\omega_2 z_2} e^{x_2 \beta_2} \Lambda_2)$$

so the two-dimensional integral factors into two one-dimensional integrals. To see why this integrand causes problems, consider the behavior of the function

$$g : z \rightarrow \exp(n\omega z - e^{\omega z} c)$$

for different values of n, ω and c .

To get a realistic picture, we used the parameters obtained in the fixed effects model to obtain values of c which correspond to the integrated intensity (for a detailed description of the model, see below). Figure 4.1 shows the integrated intensity versus the number of exacerbations for each state to illustrate which values occurred in our application. Figure 4.2 show g for a “typical” case (number of exacerbations = 4, $c = 8$) and for the extreme case (number of exacerbations = 17, $c = 3$). Each plot shows two plots for values of $\omega = 0.2$ and 1.

These plots show that the function is very different from a polynomial. It also shows that for the extreme case, different values of ω have quite a dramatic impact on the function. An increase in the number of points for a Gauss-Hermite integration will hardly improve the precision, since this spreads the abscissas over a wider range.

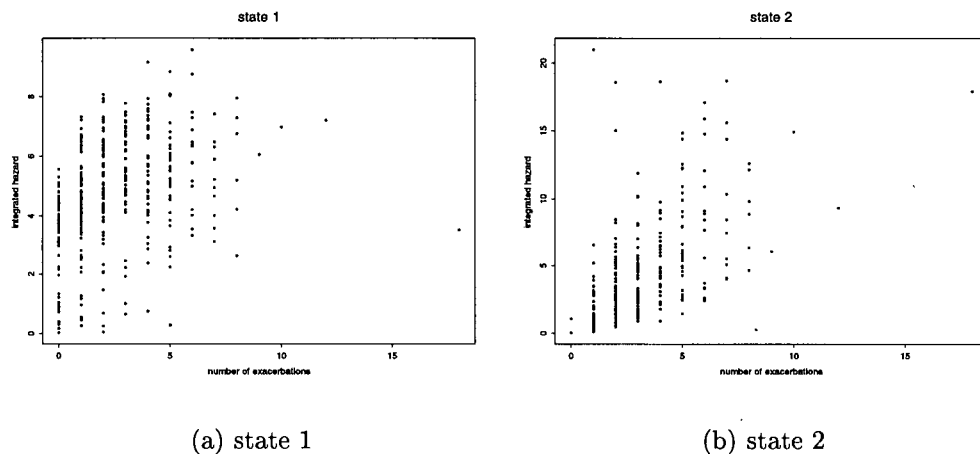


Figure 4.1: Integrated intensity versus number of exacerbations

To overcome this difficulty, we used Romberg integration to evaluate the integral. This method interpolates the function values between abscissas and works for general smooth integrands to be integrated over a finite interval. This is a suitable method in our case since we know that the integrand is almost 0 for large values. We gain more precision when using more abscissas since the additional abscissas are allocated to locations where the function is non-negligible. We set the integration limits at -4 and 4 , which corresponds for fixed effects ($\omega = 0$) to a loss of precision of roughly 10^{-5} for the integral; i.e. we could get any value in the interval 1 ± 10^{-5} when the integral equals 1.

4.7 Model Assessment

To assess how suitable the random effects models are for the data set at hand, we construct residuals based on the cumulative transition intensities

$$\Lambda_{ij}(t_{i,k-1}, t_{ik} | H_i(t_{ik}), u_{ij}, \theta_j) = \int_{t_{i,k-1}}^{t_{ik}} \lambda_{ij}(s, \theta_j | H_i(s), u_{ij}) ds$$

($j = 1, 2; i = 1, \dots, m$) over consecutive transition times. This is in analogy to the construction of residuals in survival analysis (see Lawless [15], page 281). If all the parameters including the random effects are known, these Λ_{ij} are exponentially distributed with mean 1, since the survival function for the t_{ik}

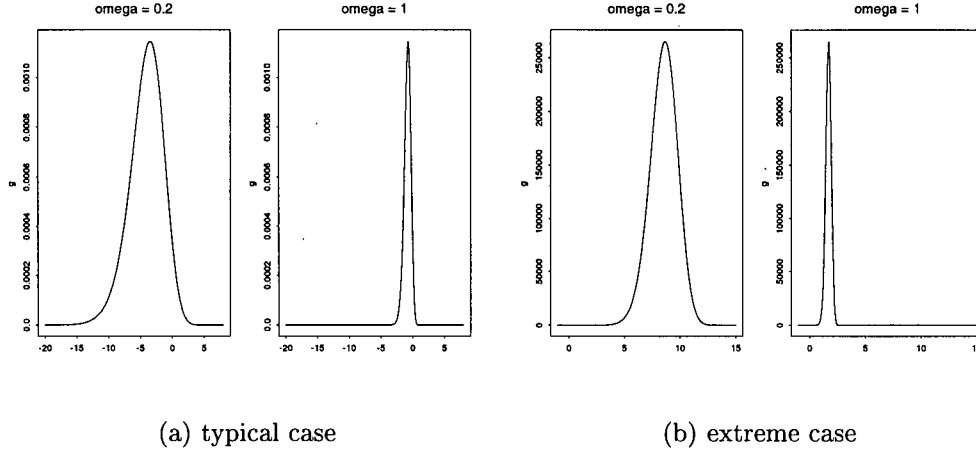


Figure 4.2: g for a typical and an extreme case (see text for definition) for different values of ω

is given by $S(t) = \exp(-\Lambda_{ij})$ and $S(t)$ is uniformly distributed.

To use this property, we have to estimate the random effects u_{ij} . In analogy to Section 2.5, we use the posterior means as estimates \hat{u}_{ij} . With this, we obtain the following residuals: If t_{ik} is not censored we take

$$\hat{r}_{ijk} = \Lambda_{ij}(t_{i,k-1}, t_{ik} | H_i(t_{ik}), \hat{u}_{ij}, \hat{\theta}_j),$$

and if t_{ik} is censored we take

$$\hat{r}_{ijk} = \Lambda_{ij}(t_{i,k-1}, t_{ik} | H_i(t_{ik}), \hat{u}_{ij}, \hat{\theta}_j) + 1.$$

This last definition is motivated by the fact that the Λ_{ij} are exponential random variables with mean 1, so for any $\Lambda > 0$, we have

$$E(\Lambda_{ij} | \Lambda_{ij} > \Lambda) = \Lambda + 1.$$

Hence to construct a residual for a censored event-time, we first evaluate a residual as if the event-time was not censored and then adjust this residual. These residuals can then be compared to a sample from an exponential distribution with a probability plot.

4.8 Results for the Alternating Renewal Approach

4.8.1 Models Fit

We fit models with different parametrizations of the effects of the time-scales with and without random effects. As pointed out earlier, we are interested in the time-scales themselves as they describe the patterns over time and in the influence of different models for the time-scales on the estimates of the covariate effects.

We model states 1 and 2 with separate sets of parameters. Initially we anticipated a parsimonious model with common parameters for the two states, but the numerical problems we had took so much time that we did not explore this further. We ended up fitting only completely separate models for the two states.

We scaled the event-times by 100 to achieve more numerical stability; the effect estimates for the time-scales in the linear parametrization are given in these units. 100 days also corresponds to roughly three months, which was the schedule for patient visits to the clinic, so is an interpretable unit of time. We experienced numerical difficulties with the parametrization of the effects of the time scales as splines, so we have no results for splines. We are not sure of the source of this difficulty.

For the piecewise parametrization of the effects of the time-scales, we need to choose intervals on the time-axis for each of the 4 scales (Markov state 1, renewal state 1, Markov state 2 and renewal state 2) where the corresponding effect is constant. A possible choice for these pieces is given by the quartiles of the corresponding sample of event-times, which are the inter-event times for the renewal scales and the times of events for the Markov scales. For convenience we chose the same pieces for the Markov scales in both states; we took the quartiles from the distribution of the end of exacerbations, since they were a bit more evenly spread. Figure 4.3 shows the distribution of the end of exacerbations as the Markov component and the backward recurrence times for both states as the renewal components. Table 4.1 lists the quartiles

time-scale	1st quartile	median	3rd quartile
Markov	191	424	701
renewal state 1	43	109	230
renewal state 2	20	35	58

Table 4.1: Quartiles for the time-scale parametrizations (in days)

for reference.

We chose the effect for the first quartile as the baseline (effect = 0), and hence the effects for the other intervals are changes in the transition intensity from this baseline. For example, the estimated effect for the renewal scale in state 1 corresponds to the exacerbation-free time periods. The effects of this renewal time-scale are allowed to be different for the different quartiles (that is, if the exacerbation-free time period lasted for 1-43, 44-109, 110-230 or 231-1095 days) and the effects for the last three quartiles describe the change in the transition intensity into exacerbation compared to the transition intensity into exacerbation if the exacerbation-free time period lasted only 1-43 days.

4.8.2 Fixed Effects

Table 4.2 presents the results for the fixed effects models with the linear and piecewise constant parametrizations of the time-scales. The estimates and 95% pointwise confidence intervals for the piecewise constant parametrization of the time-scales are given in Figures 4.4 and 4.5. For the transition into exacerbation, the estimates for the baseline covariates are almost identical for the two parametrizations of the time-scales. The estimates indicate both the low and high dose treatment are beneficial in that they significantly reduce the probability of starting an exacerbation compared to placebo, with the transition intensity for the high dose and low dose groups being roughly $\exp(-0.39) \approx 68\%$ and $\exp(-0.19) \approx 83\%$ times the transition intensity for the placebo group. Patients with a higher EDSS tend to have a higher probability to start a exacerbation and older patients tend to have a lower probability to start a exacerbation. There is some indication that female patients have a higher probability to start a exacerbation and that patients who had MS

for transition into exacerbation						
	linear time-scales			piecewise time-scales		
covariate	estimate	s.e.	z-score	estimate	s.e.	z-score
low dose	-0.1936	0.0765	-2.53	-0.1858	0.0765	-2.43
high dose	-0.3940	0.0826	-4.77	-0.3855	0.0826	-4.67
EDSS	0.0686	0.0259	2.65	0.0638	0.0259	2.46
gender	0.1008	0.0734	1.37	0.1034	0.0734	1.41
age/10	-0.1503	0.0530	-2.84	-0.1520	0.0530	-2.87
time since onset	-0.0066	0.0063	-1.05	-0.0047	0.0063	-0.75
renewal	-0.1822	0.0188	-9.67			
Markov	-0.0019	0.0111	-0.17			

for transition out of exacerbation						
	linear time-scales			piecewise time-scales		
covariate	estimate	s.e.	z-score	estimate	s.e.	z-score
low dose	0.0224	0.0763	0.29	0.0340	0.0764	0.44
high dose	0.0141	0.0825	0.17	0.0215	0.0825	0.26
EDSS	-0.0287	0.0270	-1.06	-0.0386	0.0273	-1.41
gender	-0.0113	0.0749	-0.15	-0.0010	0.0750	-0.01
age/10	-0.0974	0.0513	-1.90	-0.0955	0.0513	-1.86
time since onset	0.0126	0.0064	1.95	0.0121	0.0064	1.89
renewal	0.3276	0.0684	4.79			
Markov	-0.0335	0.0107	-3.13			
log-likelihood	-2144.51 (both transitions)			-2103.18 (both transitions)		

Table 4.2: Results for the fixed effects models

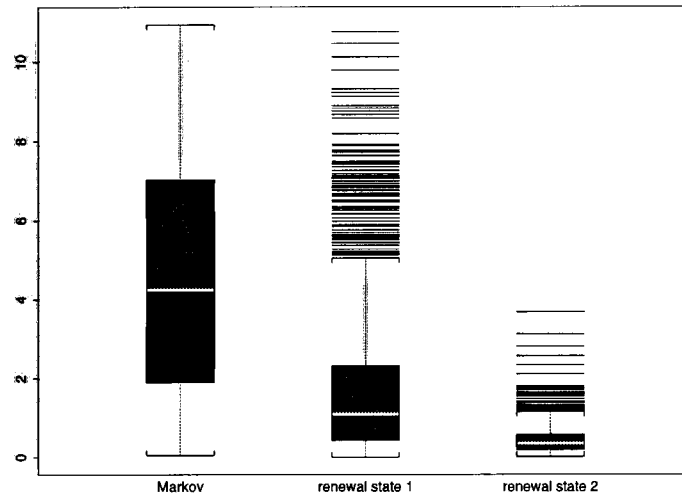


Figure 4.3: Boxplots of Markov and renewal times (in units of 100 days)

for a longer time have a decreased probability of starting a exacerbation, but neither effect is significant. It is interesting that time since onset and gender both showed a pattern in the exploratory analysis (see Figure 3.1, where the rate of exacerbations is shown versus the covariates) but are not significant, whereas there are no evident patterns for EDSS and age, both of which are estimated as being significant.

The estimated time-scales for the piecewise specification for the transition into exacerbation are plotted in Figure 4.4 (for reference, the quartiles are given in Table 4.1). The baseline probability here is the probability of starting a exacerbation after being exacerbation-free for less than 43 days. The estimated renewal component indicates that the probability of starting a exacerbation is not significantly different from baseline if the patient was exacerbation-free between 43 and 109 days. If the patient has been exacerbation-free for 110 to 230 days however, this baseline probability is decreased significantly compared to the baseline probability. The decrease from the baseline probability is even more significant if the patient was exacerbation-free for more than 231 days (with the probability of starting a exacerbation being roughly

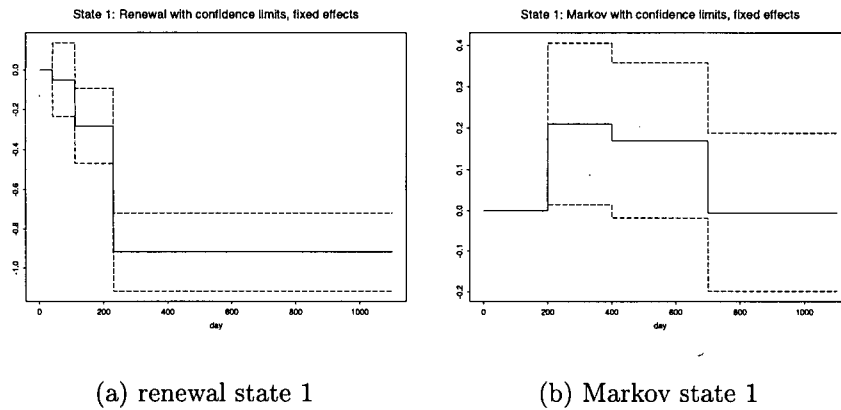


Figure 4.4: Estimated effect of time-scales for transition into exacerbation with pointwise confidence intervals, fixed effects model

$\exp(-0.9) \approx 40\%$ that of the baseline). For the Markov-component, there seems to be a slight increase of the probability to start a exacerbation for the time period between 192 and 424 days compared to the 1-191 day time period; for the other time periods there is no significant difference from the 1-191 day time period. The fact that the renewal time-scale shows a decrease in the transition into exacerbation probability is surprising, since one expects the probability of starting a exacerbation to increase the longer a patient remains exacerbation-free.

For the transition out of exacerbation, the estimated effects are slightly different depending on the parametrization of the time-scales. But this is primarily for the estimates with a low z-score, which means that they are not well-determined. The estimates for age and time since onset are almost identical for the two parametrizations, but the estimates for EDSS show some discrepancy. The only covariates that seem to have an effect on the transition out of exacerbation are time since onset, age and EDSS. Patients who had MS for a longer time tend to have a higher probability to end a exacerbation. Patients who are older or have a higher EDSS, on the other hand, tend to have lower probabilities to end a exacerbation.

Figure 4.5 shows the effects of the time-scales for the transition out of exacerbation. As expected, the renewal scale increases quite steeply (since

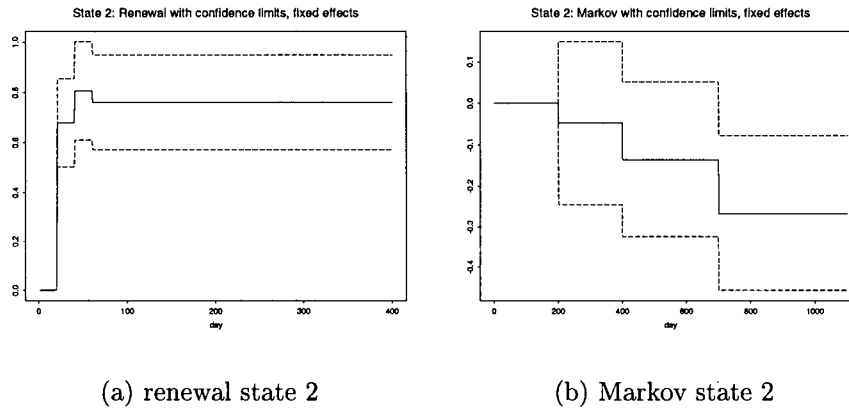


Figure 4.5: Estimated effect of time-scales for transition out of exacerbation with pointwise confidence intervals, fixed effects model

we expect the probability of ending a exacerbation to increase the longer the exacerbation lasts), but then is almost constant. This indicates that once the exacerbation lasts more than 20 days, the probability does not increase much with length any more. The Markov scale indicates that the probability to end a exacerbation decreases over time. But the decrease is modest; only the time period 702-1095 days exhibits a significant reduction compared to the time period 1-191 days.

Table 4.2 also gives the log-likelihoods for the linear and the piecewise parametrizations. These models are not nested, so we cannot use the usual χ^2 test for the differences in log-likelihood. Nevertheless, the reduction in likelihood per added parameter is still useful to judge how much the additional parameters contribute to the model fit. In this case, the gain in the maximized log-likelihood is 41.33 with an additional 8 parameters for the piecewise specification compared to the linear one. This indicates a slightly better fit of the piecewise specification.

4.8.3 Random Effects

As an initial description of the variation in the length of time in both states, we plotted the variance versus the mean length of times in one state for individual patients (Figure 4.6). The plots indicate a lot of variability in state 1, but

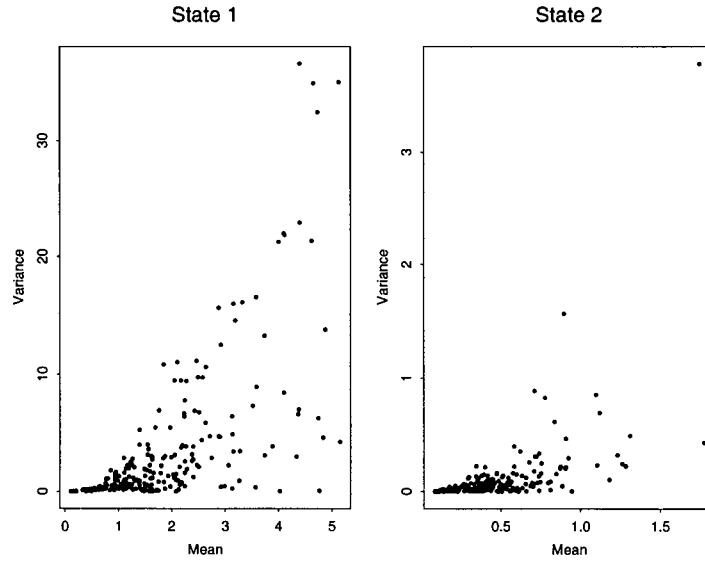


Figure 4.6: Variance versus mean for the lengths of inter-event times (in 100 days) for individual patients

not much in state 2. Hence we anticipate that the estimated variance for the random intercept in state 1 will be fairly large and the variance for the random intercept in state 2 will be modest.

We also want to assess the need to incorporate the correlation between the sojourn times in state 1 and state 2. Figure 4.7 shows the relationship between the mean sojourn time in state 1 versus the mean sojourn time in state 2. There is not much indication of a correlation between the sojourn times in the two states.

The optimization for the case of correlated random effects was very slow, so we only fit models with independent random effects (that is, $\omega_3 = 0$). We could not get the Newton-Raphson algorithm to converge for the variance parameters ω_l , $l = 1, 2$. So we evaluated profile-likelihoods for fixed values of ω_l , $l = 1, 2$ instead; that is, we obtained

$$h(\omega) = \max_{\theta_j, j=1,2} \log L(\theta, \omega).$$

Figure 4.8 shows the log-likelihood evaluated on a grid of values for

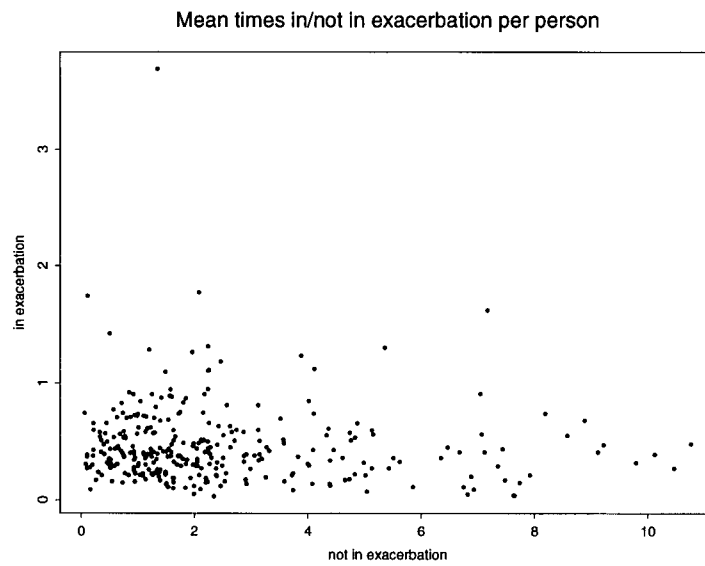


Figure 4.7: Mean sojourn time (in 100 days) in state 1 versus state 2

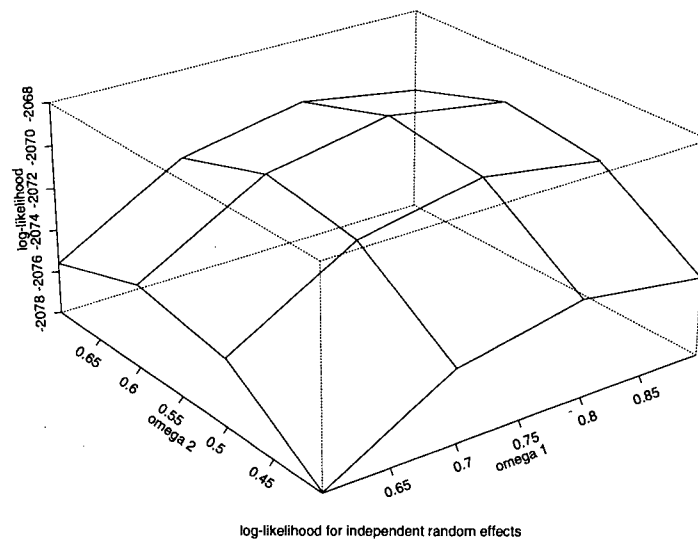


Figure 4.8: Profile log-likelihood for independent random effects

for transition out of no exacerbation						
	linear time-scales			piecewise time-scales		
covariate	estimate	s.e.	z-score	estimate	s.e.	z-score
low dose	−0.2547	0.1320	−1.93	−0.2569	0.1319	−1.95
high dose	−0.4959	0.1380	−3.59	−0.4932	0.1381	−3.57
EDSS	0.0798	0.0448	1.78	0.0783	0.0449	1.74
gender	0.1432	0.1220	1.17	0.1396	0.1221	1.14
age/10	−0.2000	0.0888	−2.25	−0.1995	0.0888	−2.25
time since onset	−0.0083	0.0105	−0.79	−0.0082	0.0105	−0.78
renewal	−0.0187	0.0212	−0.88			
Markov	−0.0403	0.0121	−3.33			

for transition out of exacerbation						
	linear time-scales			piecewise time-scales		
covariate	estimate	s.e.	z-score	estimate	s.e.	z-score
low dose	−0.0131	0.1196	−0.11	0.0162	0.1213	0.13
high dose	−0.0201	0.1264	−0.16	−0.0318	0.1273	−0.25
EDSS	−0.0262	0.0405	−0.65	−0.0504	0.0418	−1.21
gender	−0.0675	0.1160	−0.58	−0.0529	0.1166	−0.45
age/10	−0.1235	0.0859	−1.44	−0.0789	0.0796	−0.99
time since onset	0.0120	0.0100	1.20	0.0075	0.0099	0.76
renewal	0.9594	0.0791	12.13			
Markov	−0.0342	0.0123	−2.78			
log-likelihood	−2068.54 (both transitions)			−2012.01 (both transitions)		

Table 4.3: Results for the random intercepts models

(ω_1, ω_2) . From this plot we can see that the maximum on this grid is attained at $(\omega_1, \omega_2) = (0.8, 0.6)$. Hence we fit a random effects model with these values, which correspond to a standard deviation of 0.8 for the transition intensity into exacerbation and a standard deviation of 0.6 for the transition intensity out of exacerbation. We treated these values as known for all the results presented in this section.

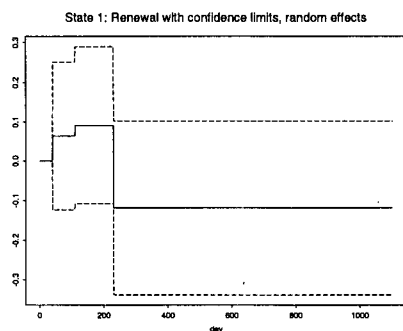
Table 4.3 shows the results for the model with random intercepts. Figures 4.9 and 4.10 show the estimated piecewise constant time-scales. For the transition out of no exacerbation, the effect estimates for the baseline covari-

ates are almost identical for the linear and piecewise constant parametrizations. There is a significant treatment effect for low dose and high dose treatment. Here the transition probability into exacerbation for a patient with high dose treatment is about $\exp(-0.49) \approx 61\%$ of the transition probability of a placebo patient with the same intercept. For a low dose patient, the transition probability is about $\exp(-0.25) \approx 78\%$ of that of a placebo patient with the same intercept. The only other baseline covariate that is significant is age, with older patients having a lower transition probability into exacerbation. Female patients seem to have a higher transition probability, but it is not significant. Similarly, higher EDSS seems to increase the transition probability, but again the effect is not significant. The estimated piecewise constant time-scale in Figure 4.9 indicates that there is no effect of the backward recurrence time on the transition probability. The estimated Markov scale shows a slight decrease over time, that is significant only in the last quartile.

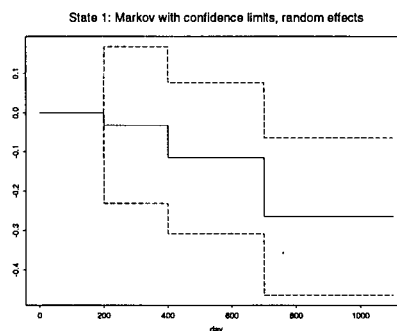
For the transition out of exacerbation we see that the estimated effects for the baseline covariates are influenced by the parametrization of the time-scales, with no consistent relationship between the effects in the two parametrizations. Overall, no effect in either parametrization is significant. There is some indication that higher EDSS and higher age lead to longer exacerbations whereas time since onset seems to contribute to shorter ones. The estimated time-scales for the piecewise parametrization shown in Figure 4.10 indicate that the renewal scale increases very sharply from baseline in the second time-interval and keeps on increasing the longer the exacerbation lasts. The Markov scale shows a decreasing trend, although the estimated effect is significantly smaller than 0 only in the last interval.

4.9 Discussion

In this section we compare the fixed effects and random effects model. Although we have discussed that the fixed effects model does not allow for correlation of the data within an individual, it is closely related to the random effects model and so it is instructive to see how the introduction of random intercepts changes the estimates. We first compare the estimated effects for the baseline covariates. We discussed already in Chapter 1 that the effect

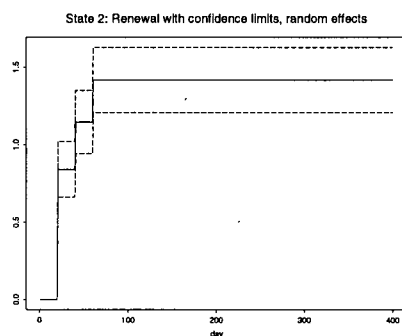


(a) renewal state 1

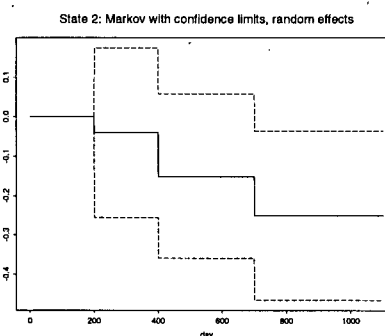


(b) Markov state 1

Figure 4.9: Estimated time-scales for transition into exacerbation with pointwise confidence interval, random intercept model



(a) renewal state 2



(b) Markov state 2

Figure 4.10: Estimated time-scales for transition out of exacerbation with pointwise confidence interval, random intercept model

covariate	fixed effects	random intercept	ratio
for transition out of no exacerbation			
low dose	-0.1858	-0.2569	0.72
high dose	-0.3855	-0.4932	0.78
EDSS	0.0638	0.0783	0.86
gender	0.1034	0.1396	0.74
age/10	-0.1520	-0.1995	0.76
time since onset	-0.0047	-0.0082	0.57
for transition out of exacerbation			
low dose	0.0340	0.0162	2.10
high dose	0.0215	-0.0318	-0.68
EDSS	-0.0386	-0.0504	0.77
gender	-0.0010	-0.0529	0.02
age/10	-0.0955	-0.0789	1.21
time since onset	0.0121	0.0075	1.61

Table 4.4: Estimated effects of baseline covariates for fixed effects and random intercept models for the model with a piecewise constant parametrization of the time-scales

estimates for the fixed and random effects models are strictly speaking not comparable, since the effect estimates for the random effects model have to be interpreted conditional on the random effects. But we have already mentioned in Chapter 2 that Diggle, Liang and Zeger ([7], p. 141) discuss a model where the ratios of the fixed to the random effects coefficients are constant (with a constant smaller than one related to the variance of the random effects). Although the alternating renewal processes are a different model, it is still interesting to compare the size of the coefficients to see if the relationship in this case is similar. Since we have seen in the last section that the linear parametrization of the time-scales were too simple to capture the renewal scale in particular, we compare the models with the piecewise parametrization only.

Table 4.4 shows the coefficients from the fixed and random effects models along with the ratio of the fixed and random effects estimates. For the transition into exacerbation, this ratio is fairly consistent and smaller than 1. This is what we anticipated in the previous paragraph. For the transition out

covariate	fixed effects	random intercept
for transition out of no exacerbation		
low dose	-2.43	-1.95
high dose	-4.67	-3.57
EDSS	2.46	1.74
gender	1.41	1.14
age/10	-2.87	-2.25
time since onset	-0.75	-0.78
for transition out of exacerbation		
low dose	0.44	0.13
high dose	0.26	-0.25
EDSS	-1.41	-1.21
gender	-0.01	-0.45
age/10	-1.86	-0.99
time since onset	1.89	0.76
log-likelihood	-2103.18	-2012.01

Table 4.5: Z-scores for different models

of exacerbation, however, this ratio is very different for different covariates; this might partly be due to the fact that the standard errors for these estimates are quite high for this transition for both the fixed and random effects models.

Table 4.5 compares the z-scores obtained from the fixed effects and random effects models. Although the estimated size of the coefficients has a different interpretation in the fixed and random effects models, the z-scores are used to judge if a covariate has an effect on the transition probabilities and are comparable in this sense. For example, the estimate for the high dose treatment has a z-score of less than -3 in both models, which leads in both cases to the conclusion that the high dose treatment is beneficial for the patients since it reduces the probability of starting an exacerbation. On the other hand, the time since onset is not estimated to be significant in either model and hence we conclude that this covariate has no influence on the transition probability into exacerbation. Compared in this way, the table shows that the conclusions are quite similar for the random effects model and the fixed effects

model. The biggest difference is EDSS, which is estimated to be significant in the fixed effects model, but not significant at the 5% level in the random effects model. The last line of Table 4.5 shows the maximized log-likelihood. The difference between the values for the log-likelihood is 91.17 with an additional 2 parameters in the random effects model. This is a considerable reduction, although we cannot use the usual χ^2 test since the extra parameters in the random effects model are the variances of the random intercepts; the fixed effects model is obtained by letting these variances go to 0, which corresponds to the boundary of the parameter space for the fixed effects model.

The estimated effects for both the renewal and Markov time-scales for state 1 are quite different for the fixed (Figure 4.4) and random effects models (Figure 4.9). The estimated effects for the renewal time-scale decrease significantly over time for the fixed effects model, whereas in the random effects model there are no clear differences over time. This suggests the renewal effect may be reflecting the fact that different patients have very different probabilities to remain exacerbation-free beyond the differences explained by the covariates. Then patients with a tendency to remain exacerbation-free for a long time would contribute to a negative effect in the renewal scale for a fixed effects model, whereas in a random effects model these patients would have a higher estimated intercept so that on an individual level the backward recurrence time is not important anymore. The estimated effects for the Markov scale for state 1 are basically 0 for the fixed effects model, and slightly decreasing in the random effects model. The fact that the effects for the Markov scale are not decreasing in the fixed effects model is a bit surprising, since we have seen in Chapter 3 that the rates have a tendency to decrease over time.

For state 2, the estimated effects for the renewal and Markov scale for the fixed (Figure 4.5) and random effects models (Figure 4.10) show qualitatively the same patterns over time. In fact, the Markov scale is almost identical. The estimated effects for the renewal scale for both models show a sharp increase for the probability to end an exacerbation for all time-periods compared to baseline. For the fixed effects model, the estimated effects are almost the same for all exacerbations that last more than 20 days, while for the random effects model the effects keep increasing over time. Hence the estimated effects for the renewal scale for state 2 suggests a dichotomy between long and short

exacerbations for the fixed effects model. The probability to end the exacerbation increases a lot after 20 days, but then remains almost the same. In the random effects model, on the other hand, the probability keeps increasing as one would expect. Again this can be explained by differences between patients who tend to have long exacerbations and patients who tend to have short exacerbations. If this difference is not explained sufficiently by the covariates, this leads to pooling short and long exacerbations in the fixed effects model and hence a fairly uniform probability of ending the exacerbation. In a random effects model, the random intercept allows for individual differences and then the renewal scale estimates the effect of the length of the exacerbation within this individual.

We noted earlier that our choice of the bivariate normal distribution for the random effects is different from the usual choice in models for repeated failure times of a single type. Such random effects are often assumed to be Gamma distributed. This simplifies the form of the marginal likelihood; for a detailed discussion, see Lawless [16], Venables and Ripley [32] p. 242 or Aalen and Husebye [1]. However, bivariate extensions of the Gamma distribution allow only positive correlations and are hence not a natural choice for our model.

Chapter 5

A Marked Point Process Model

5.1 Introduction

In this chapter, we introduce another point process model. In the previous chapter, we have seen how we can use alternating renewal processes to model the MS disease process. This leads to a plausible and flexible stochastic model. However, as discussed in Chapter 1, the interpretation of treatment effects and baseline covariates is easier in a model that models the mean exacerbation counts and mean exacerbation lengths directly. The marked point process model introduced in this chapter accomplishes this.

We model each individual as a marked point process, for which we need to model the process of events K_t , which is given by the sequence of days an exacerbation started, and the process of marks M_t , which is given by the lengths of exacerbations. Although Lawless [17] indicates that this approach can be developed in a continuous time setting, for convenience we considered only a discrete time setting. We observe the underlying continuous process (K_t, M_t) at a set of fixed and equally-spaced time-points t_1, \dots, t_k . We give more detail below.

We have seen in Chapter 3 that a model for the marginal means does not specify a point process completely. To avoid making strong assumptions on the process, we use an approach to inference that does not need a full distributional model for the point process: the generalized estimating equations (GEE) approach. In summary, the GEE approach, as often implemented,

proceeds according to the following steps:

- derive estimating equations obtained under independence and full distributional assumptions
- study the asymptotic behaviour of estimators arising as solutions of these estimating equations under weaker distributional assumptions, typically specifications relating only to the marginal distributions rather than the joint distributions
- estimate this asymptotic covariance matrix.

We use a Taylor-expansion of the score vector about the true value to derive the asymptotic covariance matrix of the estimates under the reduced assumptions.

In this chapter, we first derive score equations for Poisson processes and then show how the GEE approach can use these equations to obtain an inference valid for more general point processes. We then extend this approach to the setting of marked point processes.

Let us first introduce some notation. Let $K_i(t) = K(t, \mathbf{x}_i)$ be independent processes for $i = 1, \dots, m$, observed at a fixed set of time points $0 < t_1 < \dots < t_{k_i} < T$. Here T is the end of the observation period and m is the number of individuals. We assume that these time points are equally-spaced, since in our application we have daily data on individuals over a span of three years and we can choose our time points to be equally-spaced. All individuals are observed at the same time points unless the individual drops out before time T , which leads to a different number k_i of observations on this individual. The number of events for individual i in $[0, t]$ is denoted by $K_i(t)$. We model the counts $n_i(t_l) = K_i(t_l) - K_i(t_{l-1})$ occurring in each interval $(t_{l-1}, t_l]$, $l = 1, \dots, k_i$, so that $n_i(t_l)$ counts the number of exacerbations starting in this interval. We abuse notation and denote both the random variables and observed counts by $n_i(t_l)$ depending on the context. Note that the counts $\{n_i(t_l); l = 1, \dots, k_i\}$ on an individual are independent if the K_i are Poisson processes.

The equal spacing of the time points at which individuals are observed simplifies the modeling, since this way we don't have to take the lengths of the

intervals $(t_{l-1}, t_l]$ into account. The dependence on t in our context reflects the changes in the distribution of $n_i(t)$ over time. We incorporate the covariate vector \mathbf{x}_i into the specification for the mean counts. Let us denote the mean for these counts by

$$E(n_i(t)) = g(\mathbf{x}_i, t; \boldsymbol{\theta}) =: g_i(t, \boldsymbol{\theta}),$$

where we suppress the \mathbf{x}_i for notational convenience.

5.2 Estimating Equations for Poisson Processes

Under specification of the joint distribution, maximum likelihood estimates are obtained by solving estimating equations derived from the derivative of the likelihood. The generalized estimating equations approach uses these same estimating equations, but drops the assumption that these equations are score equations. In this section we derive score equations under the assumption that $K_i(t)$ are independent Poisson processes. In the next section we study the properties of the solutions of these equations after dropping the assumption that the data within individuals are independent.

In this section we assume that:

1. data from different individuals are independent
2. $E(n_i(t)) = g_i(t, \boldsymbol{\theta})$
3. $n_i(t)$ is a Poisson random variable
4. $n_i(t_j)$ and $n_i(t_l)$ are independent for $j \neq l$.

We refer to these assumptions subsequently as the **Poisson assumptions**. Assumptions 3. and 4. are satisfied if the processes $K_i(t, \mathbf{x}_i)$ are Poisson processes.

Under these assumptions, the likelihood would be

$$L(\boldsymbol{\theta}; \{n_i(t_l)\}) = \prod_{i=1}^m \prod_{l=1}^{k_i} \frac{e^{-g_i(t_l, \boldsymbol{\theta})} g_i(t_l, \boldsymbol{\theta})^{n_i(t_l)}}{n_i(t_l)!},$$

with log-likelihood

$$\log L(\boldsymbol{\theta}) = \text{const} + \sum_{i=1}^m \sum_{l=1}^{k_i} \{n_i(t_l) \log g_i(t_l, \boldsymbol{\theta}) - g_i(t_l, \boldsymbol{\theta})\}.$$

This leads to the score vector

$$\mathbf{U}_1(\boldsymbol{\theta}) = \frac{\partial \log L}{\partial \boldsymbol{\theta}} = \sum_{i=1}^m \sum_{l=1}^{k_i} \left\{ \frac{n_i(t_l) - g_i(t_l, \boldsymbol{\theta})}{g_i(t_l, \boldsymbol{\theta})} \right\} \frac{\partial g_i(t_l, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \quad (5.1)$$

Let $\hat{\boldsymbol{\theta}}$ be the solution to $\mathbf{U}_1(\boldsymbol{\theta}) = \mathbf{0}$. We assume the solution is unique and maximizes the log-likelihood. Under the Poisson assumptions $\hat{\boldsymbol{\theta}}$ is the MLE for $\boldsymbol{\theta}$. In the next section we relax the assumptions and show that we can still use $\hat{\boldsymbol{\theta}}$ for inference.

5.3 Generalized Estimating Equations for Point Processes

Lawless and Nadeau [18] use the score equations obtained under the Poisson assumptions and then weaken the assumptions to generalize their approach to a point process setting. We review their approach here. We drop the independence assumption for data within individuals and we no longer assume that the counts have a Poisson distribution. More precisely, we simply assume that

1. data from different individuals are independent
2. $E(n_i(t)) = g_i(t, \boldsymbol{\theta})$.

Since we are only sketching the argument, we do not list the necessary regularity conditions here.

The basic idea is to use the estimating equations obtained under the Poisson assumptions as in the previous section to get an estimate and use the so-called sandwich estimator for the asymptotic covariance matrix of that estimate. The next paragraph sketches the argument. The assumption $E(n_i(t)) = g_i(t, \boldsymbol{\theta})$ ensures that the score equations are unbiased, that is $E(\mathbf{U}_1(\boldsymbol{\theta})) = \mathbf{0}$.

This is important since, combined with the independence across individuals, it leads to a consistent estimator, that is $\hat{\theta} \rightarrow \theta$ in probability as $m \rightarrow \infty$.

Let $\hat{\theta}$ be the solution to $U_1(\theta) = 0$ as in the previous section. The asymptotic distribution of $\hat{\theta}$ is found by using a Taylor-expansion of U_1 about θ :

$$0 = U_1(\hat{\theta}) = U_1(\theta) + \frac{\partial}{\partial \theta} U_1(\theta^*)(\hat{\theta} - \theta),$$

where θ^* is between $\hat{\theta}$ and θ . Since $\hat{\theta}$ is consistent, $\theta^* \rightarrow \theta$ in probability and we can substitute θ for θ^* in all the limiting expressions below. The Taylor-expansion can be written as

$$\sqrt{m}(\hat{\theta} - \theta) = - \left(\frac{1}{m} \frac{\partial}{\partial \theta} U_1(\theta) \right)^{-1} \frac{1}{\sqrt{m}} U_1(\theta).$$

Because U_1 is a sum of m independent random variables, each with mean 0, we can apply a Central Limit Theorem. Setting (assuming the limit exists and is finite)

$$I_0(\theta) = \lim_{m \rightarrow \infty} \frac{1}{m} \text{Var}(U_1(\theta)),$$

we thus have

$$\sqrt{m} \frac{1}{m} U_1(\theta) \xrightarrow[m \rightarrow \infty]{d} N(0, I_0(\theta)),$$

Hence, denoting

$$I_1(\theta) = - \lim_{m \rightarrow \infty} E \left[\frac{1}{m} \frac{\partial}{\partial \theta} U_1(\theta) \right],$$

we get

$$\sqrt{m}(\hat{\theta} - \theta) \xrightarrow[m \rightarrow \infty]{d} N(0, I_1^{-1}(\theta) I_0(\theta) I_1^{-1}(\theta)'). \quad (*)$$

Under the Poisson assumptions, it can be shown that $I_0 = I_1$ (see (5.3) and (5.4) below); so we obtain the well-known result for the asymptotic distribution of the MLE.

In practice this result would be utilized in the form

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \frac{1}{m} \mathbf{I}_1^{-1}(\boldsymbol{\theta}) \mathbf{I}_0(\boldsymbol{\theta}) \mathbf{I}_1^{-1}(\boldsymbol{\theta})')$$

where with abuse of notation

$$\mathbf{I}_1(\boldsymbol{\theta}) = -E \left[\frac{1}{m} \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{U}_1(\boldsymbol{\theta}) \right]$$

and

$$\mathbf{I}_0(\boldsymbol{\theta}) = \frac{1}{m} \text{Var}(\mathbf{U}_1(\boldsymbol{\theta})).$$

The asymptotic distribution of $\hat{\boldsymbol{\theta}}$ depends on $\mathbf{I}_0(\boldsymbol{\theta})$ and $\mathbf{I}_1(\boldsymbol{\theta})$. To enable us to carry out inference for $\boldsymbol{\theta}$, we have to provide estimates for these matrices. This is done in the remainder of this section.

Note that, since data from different individuals are independent,

$$\begin{aligned} & \frac{1}{m} \text{Var}(\mathbf{U}_1(\boldsymbol{\theta})) \\ &= \frac{1}{m} \sum_{i=1}^m \text{Var} \left[\sum_{l=1}^{k_i} \left\{ \frac{n_i(t_l) - g_i(t_l, \boldsymbol{\theta})}{g_i(t_l, \boldsymbol{\theta})} \right\} \frac{\partial g_i(t_l, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] \\ &= \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^{k_i} \sum_{l=1}^{k_i} \frac{1}{g_i(t_j, \boldsymbol{\theta})} \frac{1}{g_i(t_l, \boldsymbol{\theta})} \frac{\partial g_i(t_j, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial g_i(t_l, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}' \text{Cov}(n_i(t_j), n_i(t_l)) \\ &= \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^{k_i} \sum_{l=1}^{k_i} \frac{\partial}{\partial \boldsymbol{\theta}} \log g_i(t_j, \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}} \log g_i(t_l, \boldsymbol{\theta})' \text{Cov}(n_i(t_j), n_i(t_l)). \end{aligned}$$

To estimate $\text{Cov}(n_i(t_j), n_i(t_l))$, we simply use the empirical covariance given by

$$\widehat{\text{Cov}}(n_i(t_j), n_i(t_l)) = (n_i(t_j) - g_i(t_j, \boldsymbol{\theta}))(n_i(t_l) - g_i(t_l, \boldsymbol{\theta})). \quad (5.2)$$

Substituting $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$, we obtain

$$\begin{aligned} \widehat{\mathbf{I}}_0(\hat{\boldsymbol{\theta}}) &= \frac{1}{m} \widehat{\text{Var}}(\mathbf{U}_1(\boldsymbol{\theta})) = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^{k_i} \sum_{l=1}^{k_i} \frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \log g_i(t_j, \hat{\boldsymbol{\theta}}) \\ & \quad \frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \log g_i(t_l, \hat{\boldsymbol{\theta}})' (n_i(t_j) - g_i(t_j, \hat{\boldsymbol{\theta}}))(n_i(t_l) - g_i(t_l, \hat{\boldsymbol{\theta}})). \end{aligned} \quad (5.3)$$

For $\mathbf{I}_1(\boldsymbol{\theta})$, we obtain (after simplification arising from $E(n_i(t_j) - g_i(t_j, \boldsymbol{\theta})) = 0$)

$$\mathbf{I}_1(\boldsymbol{\theta}) = -E \left[\frac{1}{m} \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{U}_1(\boldsymbol{\theta}) \right] = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^{k_i} \frac{\partial}{\partial \boldsymbol{\theta}} \log g_i(t_j, \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}} \log g_i(t_j, \boldsymbol{\theta})' g_i(t_j, \boldsymbol{\theta}) \quad (5.4)$$

which can be estimated by plugging in $\hat{\boldsymbol{\theta}}$.

The empirical covariance (5.2) would be a poor estimator of the covariance, since it uses only two observations. What makes this approach work is the fact that this estimator appears in the middle layer of the sandwich estimator $\mathbf{I}_1^{-1} \mathbf{I}_0 \mathbf{I}_1^{-1}$ so that these estimates for the covariance are averaged across individuals (see Diggle, Liang and Zeger [7], p.71).

5.4 Estimating Equations for Marked Poisson Processes

We now extend the Lawless and Nadeau approach to a specific case of a marked process. This allows us to incorporate the length of exacerbations into the analysis. We first introduce some notation. Let us assume that the length of an exacerbation for individual i is a random variable with mean $h_i(t, \boldsymbol{\xi}) = h(t, \boldsymbol{\xi}, \mathbf{x}_i)$ where \mathbf{x}_i is the vector of covariates. The use of time as a covariate here represents the changes in the distribution of the length of exacerbations over time. We model the total length of exacerbations $m_i(t_l)$ beginning in an interval $(t_{l-1}, t_l]$, given the number of exacerbations $n_i(t_l)$ starting in this interval (note that $m_i(t_l)$ includes the complete length of an exacerbation starting in this time interval, even if this exacerbation lasts past time t_l). We model the effects of the covariates on the counts and the effects of the covariates on the lengths of exacerbations with separate sets of parameters. Since the expected total length of exacerbations depends on the expected counts and the expected length of an exacerbation, the parameters $\boldsymbol{\xi}$ model the additional information from the lengths of exacerbations beyond that from the number of exacerbations.

We first derive estimating equations under independence assumptions. In particular, we assume that

1. data from different individuals are independent
2. $E(n_i(t)) = g_i(t, \theta)$
3. $n_i(t)$ is a Poisson random variable
4. $n_i(t_j)$ and $n_i(t_l)$ are independent for $j \neq l$.

We further assume that

5. conditional on the number of events $n_i(t_l)$, $m_i(t_l)$ is the sum of $n_i(t_l)$ independent and identically distributed random variables $X_j^{i,l}$ where $X_j^{i,l}$ has mean $h_i(t_l, \xi)$
6. $X_j^{i,l}$ is a Poisson random variable (other distributional forms could be considered, but we do not pursue this here)
7. the pairs $(n_i(t_l), m_i(t_l))$, $l = 1, \dots, k$ are independent for each $i = 1, \dots, m$.

It might be possible to weaken these assumptions, but we did not pursue this here. These assumptions imply that given $n_i(t_l)$, $m_i(t_l)$ has a Poisson distribution with mean $n_i(t_l)h_i(t_l, \xi)$. It follows that

$$E(m_i(t_l)) = g_i(t_l, \theta)h_i(t_l, \xi) \quad (5.5)$$

$$\begin{aligned} \text{Var}(m_i(t_l)) &= E(\text{Var}(m_i(t_l)|n_i(t_l))) + \text{Var}(E(m_i(t_l)|n_i(t_l))) \\ &= g_i(t_l, \theta)h_i(t_l, \xi) + g_i(t_l, \theta)h_i(t_l, \xi)^2. \end{aligned} \quad (5.6)$$

The expression for the variance is not used below and is given only to emphasize that the $m_i(t_l)$ are not Poisson random variables.

To derive the likelihood, first note

$$\begin{aligned} P(n_i(t_l), m_i(t_l)) &= P(n_i(t_l))P(m_i(t_l)|n_i(t_l)) \\ &= \frac{g_i(t_l, \theta)^{n_i(t_l)}}{n_i(t_l)!} e^{-g_i(t_l, \theta)} \frac{(h_i(t_l, \xi)n_i(t_l))^{m_i(t_l)}}{m_i(t_l)!} e^{-h_i(t_l, \xi)n_i(t_l)}. \end{aligned}$$

Hence the likelihood becomes

$$\begin{aligned}
L(\boldsymbol{\theta}, \boldsymbol{\xi}) &= \prod_{i=1}^m P(\{n_i(t_l), m_i(t_l); l = 1, \dots, k\}) \text{ by independence of individuals} \\
&= \prod_{i=1}^m \prod_{l=1}^{k_i} P(n_i(t_l), m_i(t_l)) \text{ by assumption 7} \\
&= \prod_{i=1}^m \prod_{l=1}^{k_i} \frac{g_i(t_l, \boldsymbol{\theta})^{n_i(t_l)}}{n_i(t_l)!} e^{-g_i(t_l, \boldsymbol{\theta})} \frac{(h_i(t_l, \boldsymbol{\xi}) n_i(t_l))^{m_i(t_l)}}{m_i(t_l)!} e^{-h_i(t_l, \boldsymbol{\xi}) n_i(t_l)}.
\end{aligned}$$

The log-likelihood

$$\begin{aligned}
\log L(\boldsymbol{\theta}, \boldsymbol{\xi}) &= \text{const} + \sum_{i=1}^m \sum_{l=1}^{k_i} \{n_i(t_l) \log g_i(t_l, \boldsymbol{\theta}) - g_i(t_l, \boldsymbol{\theta})\} \\
&\quad + \sum_{i=1}^m \sum_{l=1}^{k_i} [m_i(t_l) \log \{h_i(t_l, \boldsymbol{\xi}) n_i(t_l)\} - h_i(t_l, \boldsymbol{\xi}) n_i(t_l)]
\end{aligned}$$

leads to the score vectors:

$$\begin{aligned}
\mathbf{U}_1(\boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\theta}} \log L(\boldsymbol{\theta}, \boldsymbol{\xi}) = \sum_{i=1}^m \sum_{l=1}^{k_i} \left\{ \frac{n_i(t_l) - g_i(t_l, \boldsymbol{\theta})}{g_i(t_l, \boldsymbol{\theta})} \right\} \frac{\partial}{\partial \boldsymbol{\theta}} g_i(t_l, \boldsymbol{\theta}), \\
\mathbf{U}_2(\boldsymbol{\xi}) &= \frac{\partial}{\partial \boldsymbol{\xi}} \log L(\boldsymbol{\theta}, \boldsymbol{\xi}) = \sum_{i=1}^m \sum_{l=1}^{k_i} \left\{ \frac{m_i(t_l) - n_i(t_l) h_i(t_l, \boldsymbol{\xi})}{h_i(t_l, \boldsymbol{\xi})} \right\} \frac{\partial}{\partial \boldsymbol{\xi}} h_i(t_l, \boldsymbol{\xi}).
\end{aligned}$$

Note that $\mathbf{U}_1(\boldsymbol{\theta}, \boldsymbol{\xi})$ does not depend on $\boldsymbol{\xi}$ and $\mathbf{U}_2(\boldsymbol{\theta}, \boldsymbol{\xi})$ does not depend on $\boldsymbol{\theta}$. This implies that the score equations for $\boldsymbol{\theta}$ are the same as before and we have

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\xi}} \mathbf{U}_1(\boldsymbol{\theta}) &= \mathbf{0}, \\
\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{U}_2(\boldsymbol{\xi}) &= \mathbf{0}.
\end{aligned}$$

The key point is that again we have unbiased score equations:

$$E(\mathbf{U}_1(\boldsymbol{\theta})) = E(\mathbf{U}_2(\boldsymbol{\xi})) = \mathbf{0}.$$

We denote the solutions to the estimating equations $\mathbf{U}_1(\boldsymbol{\theta}) = \mathbf{0}$ and $\mathbf{U}_2(\boldsymbol{\xi}) = \mathbf{0}$ by $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\xi}}$. We assume that these solutions are unique and maximize the log-likelihood.

5.5 GEE for Marked Point Processes

As in Section 5.3, we now use the estimating equations obtained above as a basis for inference in a more general context. We drop the distributional assumptions (3) and (6) and the assumptions (4) and (7) of independence of the data within individuals. In particular, we now assume that

1. data from different individuals are independent
2. $E(n_i(t)) = g_i(t, \boldsymbol{\theta})$

and

3. conditional on the number of events $n_i(t_l)$, $m_i(t_l)$ has mean $n_i(t_l)h_i(t_l, \boldsymbol{\xi})$.

To develop the GEE approach under these assumptions, we follow the same steps as in Section 5.3. The Taylor-expansion of the score vector about $(\boldsymbol{\theta}, \boldsymbol{\xi})$ yields

$$\mathbf{0} = \begin{pmatrix} \mathbf{U}_1(\hat{\boldsymbol{\theta}}) \\ \mathbf{U}_2(\hat{\boldsymbol{\xi}}) \end{pmatrix} = \begin{pmatrix} \mathbf{U}_1(\boldsymbol{\theta}) \\ \mathbf{U}_2(\boldsymbol{\xi}) \end{pmatrix} + \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{U}_1(\boldsymbol{\theta}^*) & \mathbf{0} \\ \mathbf{0} & \frac{\partial}{\partial \boldsymbol{\xi}} \mathbf{U}_2(\boldsymbol{\xi}^*) \end{pmatrix} \begin{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\theta}} \\ \hat{\boldsymbol{\xi}} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\xi} \end{pmatrix} \end{pmatrix}$$

where $\boldsymbol{\theta}^*$ is between $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\xi}^*$ is between $\boldsymbol{\xi}$ and $\hat{\boldsymbol{\xi}}$. Because we assume that both estimators are consistent, we can replace $\boldsymbol{\theta}^*$ by $\boldsymbol{\theta}$ and $\boldsymbol{\xi}^*$ by $\boldsymbol{\xi}$ in the limiting expressions below.

Analogous to the previous derivation of the asymptotic distribution in Section 5.2 we obtain

$$\sqrt{m} \begin{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\theta}} \\ \hat{\boldsymbol{\xi}} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\xi} \end{pmatrix} \end{pmatrix} \xrightarrow[m \rightarrow \infty]{d} N \left(\mathbf{0}, \tilde{\mathbf{I}}_1^{-1} \tilde{\mathbf{I}}_0 (\tilde{\mathbf{I}}_1^{-1})' \right)$$

where

$$\tilde{\mathbf{I}}_0 = \lim_{m \rightarrow \infty} \frac{1}{m} \text{Var} \begin{pmatrix} \mathbf{U}_1(\boldsymbol{\theta}) \\ \mathbf{U}_2(\boldsymbol{\xi}) \end{pmatrix}$$

and

$$\tilde{\mathbf{I}}_1 = - \lim_{m \rightarrow \infty} E \frac{1}{m} \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{U}_1(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \frac{\partial}{\partial \boldsymbol{\xi}} \mathbf{U}_2(\boldsymbol{\xi}) \end{pmatrix}$$

(where we assume that the limits exist). We suppress the dependence of $\tilde{\mathbf{I}}_0$ and $\tilde{\mathbf{I}}_1$ on $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ for notational convenience.

The block-diagonal structure of $\tilde{\mathbf{I}}_1$ has important implications, since it means that the covariance between \mathbf{U}_1 and \mathbf{U}_2 does not enter the asymptotic covariance matrix of $\boldsymbol{\theta}$ or the asymptotic covariance matrix of $\boldsymbol{\xi}$ which are given by the corresponding matrices along the diagonal of $\tilde{\mathbf{I}}_1^{-1}\tilde{\mathbf{I}}_0(\tilde{\mathbf{I}}_1^{-1})'$. This implies that the inference for $\boldsymbol{\theta}$ is not affected by incorporating the length of exacerbations through the parameters $\boldsymbol{\xi}$. The off-diagonal matrices of $\tilde{\mathbf{I}}_1^{-1}\tilde{\mathbf{I}}_0(\tilde{\mathbf{I}}_1^{-1})'$ can also be easily calculated. They would be needed for inference concerning functions involving both θ_j and ξ_l . Since we do not pursue this here, we omit these expressions.

To estimate $Var(\mathbf{U}_1(\boldsymbol{\theta}))$, we use (5.3). For $Var(\mathbf{U}_2(\boldsymbol{\xi}))$, we obtain similarly

$$Var(\mathbf{U}_2(\boldsymbol{\xi})) = \sum_i \sum_j \sum_l \frac{\partial}{\partial \boldsymbol{\xi}} \log h_i(t_j, \boldsymbol{\xi}) \frac{\partial}{\partial \boldsymbol{\xi}} \log h_i(t_l, \boldsymbol{\xi})' \\ Cov(\{m_i(t_j) - h_i(t_j, \boldsymbol{\xi})n_i(t_j)\}, \{m_i(t_l) - h_i(t_l, \boldsymbol{\xi})n_i(t_l)\}).$$

To obtain an estimator for $Var(\mathbf{U}_2)$ we again use the empirical covariance and then replace $\boldsymbol{\xi}$ by $\hat{\boldsymbol{\xi}}$ in the resulting expression.

To obtain an estimator for $(\tilde{\mathbf{I}}_1)$, we can use (5.4) for $E(\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{U}_1(\boldsymbol{\theta}))$. For $E(\frac{\partial}{\partial \boldsymbol{\xi}} \mathbf{U}_2(\boldsymbol{\xi}))$, we obtain similarly (after simplification arising from $E(m_i(t) - n_i(t)h_i(t, \boldsymbol{\xi})) = 0$)

$$E\left(\frac{\partial}{\partial \boldsymbol{\xi}} \mathbf{U}_2(\boldsymbol{\xi})\right) = - \sum_i \sum_j \frac{\partial}{\partial \boldsymbol{\xi}} \log h_i(t_j, \boldsymbol{\xi}) \left(\frac{\partial}{\partial \boldsymbol{\xi}} \log h_i(t_j, \boldsymbol{\xi})\right)' g_i(t_l, \boldsymbol{\theta}) h_i(t_l, \boldsymbol{\xi})$$

To obtain an estimator, we plug in $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\xi}}$.

5.6 Parametrization of the Mean Functions

We now specify the form of the mean functions for the counts of exacerbations, $g(\mathbf{x}_i, t, \boldsymbol{\theta})$, and the lengths of exacerbations, $h(\mathbf{x}_i, t, \boldsymbol{\xi})$. We parameterize

$g(\mathbf{x}_i, t, \boldsymbol{\theta})$ as the product of a factor $g_0(t, \boldsymbol{\alpha})$ that is common for all individuals and incorporates the changes in the distribution of the counts over time and a factor $\exp(\mathbf{x}_i' \boldsymbol{\beta})$ that expresses the dependence of the counts on the covariates:

$$g(\mathbf{x}_i, t, \boldsymbol{\theta}) = g(\mathbf{x}_i, t; \boldsymbol{\alpha}, \boldsymbol{\beta}) = g_0(t, \boldsymbol{\alpha}) \exp(\mathbf{x}_i' \boldsymbol{\beta}).$$

For the parameterization of $g_0(t, \boldsymbol{\alpha})$ we can choose a convenient and flexible Weibull form; that is

$$g_0(t, \boldsymbol{\alpha}) = \alpha_1 t^{\alpha_2}.$$

This has the advantage that it leads to a log-linear model, since

$$E(n_i(t_l)) = g(\mathbf{x}_i, t_l; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \exp(\log \alpha_1 + \alpha_2 \log t_l + \mathbf{x}_i' \boldsymbol{\beta}).$$

Since we use estimates obtained by solving the usual score equations (obtained under independence assumptions), we can use generalized linear models software to get $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$.

Similarly, we parameterize $h(\mathbf{x}_i, t, \boldsymbol{\xi})$ as

$$h(\mathbf{x}_i, t, \boldsymbol{\xi}) = h_0(t, \boldsymbol{\gamma}) \exp(\mathbf{x}_i' \boldsymbol{\delta})$$

with

$$h_0(t, \boldsymbol{\gamma}) = \gamma_1 t^{\gamma_2}.$$

We solve $\mathbf{U}_2(\boldsymbol{\xi}) = 0$ with a Newton-Raphson algorithm to obtain $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\delta}}$.

The estimator for the asymptotic covariance matrix can be calculated directly once we have obtained $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\delta}}$. For these calculations we note that the log-linear form for h and g makes them quite convenient, because the expressions we have to evaluate involve derivatives of the logarithms of h and g .

The regression coefficients $\boldsymbol{\beta}$ describe the impact of the covariates on the mean number of exacerbations. The regression coefficients $\boldsymbol{\delta}$, on the other hand, describe the effect of the covariates on the mean length of exacerbations. We modeled these two regressions with separate sets of parameters.

With the methods described above, we can obtain two kinds of standard errors. The so-called “naive” standard errors are based solely on the information matrix \mathbf{I}_1 and are valid only if the data on each individual actually follows a marked Poisson process with Poisson distributed marks. The “robust” standard errors are obtained from the sandwich estimator and are generally valid, as discussed in the previous section.

We also evaluated standard errors for the estimated parameters using the jackknife method. We describe how we applied the jackknife here; for a more general description see [28], p.437. We remove the data on individual i and recalculate estimates of the parameters, denoted by $(\hat{\theta}^{(i)}, \hat{\xi}^{(i)})$. We then calculate the so-called pseudo-values $\theta^i = m\hat{\theta} - (m-1)\hat{\theta}^{(i)}$ and $\xi^i = m\hat{\xi} - (m-1)\hat{\xi}^{(i)}$, where m is the number of individuals. We then estimate the covariance matrix for $\begin{pmatrix} \hat{\theta} \\ \hat{\xi} \end{pmatrix}$ as

$$\frac{1}{m(m-1)} \sum_{i=1}^m \left\{ \begin{pmatrix} \hat{\theta}^i \\ \hat{\xi}^i \end{pmatrix} - \begin{pmatrix} \bar{\theta} \\ \bar{\xi} \end{pmatrix} \right\} \left\{ \begin{pmatrix} \hat{\theta}^i \\ \hat{\xi}^i \end{pmatrix} - \begin{pmatrix} \bar{\theta} \\ \bar{\xi} \end{pmatrix} \right\}',$$

where $\bar{\theta} = \frac{1}{m} \sum_{i=1}^m \theta^i$ and $\bar{\xi} = \frac{1}{m} \sum_{i=1}^m \xi^i$. Viewed in a different way, we calculate the estimates θ^i and ξ^i by eliminating the contribution of the i th individual to the vectors \mathbf{U}_1 and \mathbf{U}_2 .

We expect the jackknife method to work well for the estimation of standard errors in situations where the parameter is obtained as the solution of score equations, if the individual contributions to the score vector can be viewed as a sample of independent and identically distributed random variables. To link this to our application, let us rewrite $\mathbf{U}_1(\boldsymbol{\theta}) = \sum_i a_i(\{n_i\}, \mathbf{x}_i, \boldsymbol{\theta})$ and $\mathbf{U}_2(\boldsymbol{\xi}) = \sum_i b_i(\{n_i, m_i\}, \mathbf{x}_i, \boldsymbol{\xi})$ where $\{n_i\}$ are all event counts on individual i and $\{n_i, m_i\}$ is the set of all event counts and marks on individual i . Hence one expects the jackknife method to work well if the individual contributions a_i and b_i to the vectors \mathbf{U}_1 and \mathbf{U}_2 are independent (as they are) and roughly identically distributed. While this is appropriate for \mathbf{U}_1 , \mathbf{U}_2 involves both the counts n_i and the marks m_i , so that b_i involves a sample from a bivariate distribution. We have seen that the differences in mean time in exacerbation are quite extreme across individuals, so we don't expect the b_i s

covariate	estimate	robust s.e.	robust z-score
low dose	-0.1582	0.1089	-1.45
high dose	-0.4038	0.1120	-3.60
age/30	-0.4295	0.2114	-2.03
sex	0.1338	0.1076	1.24
time since onset	-0.0043	0.0087	-0.50
EDSS	0.0498	0.0316	1.57
time	-0.3639	0.0427	-8.52

Table 5.1: Estimated effects for the mean counts of exacerbations

covariate	naive s.e.	robust s.e.	jackknifed s.e.	naive/robust
low dose	0.0756	0.1089	0.1117	0.69
high dose	0.0811	0.1120	0.1144	0.72
age/30	0.1524	0.2114	0.2165	0.72
sex	0.0728	0.1076	0.1106	0.68
time since onset	0.0062	0.0087	0.0090	0.71
EDSS	0.0252	0.0316	0.0324	0.80
time	0.0410	0.0427	0.0429	0.96

Table 5.2: Estimated standard errors for the mean counts of exacerbations

to be identically distributed. In summary we expect the jackknife method to work well for the standard errors of $\hat{\theta}$ but it is not clear the method yields appropriate standard errors for $\hat{\xi}$.

5.7 Application to the MS Dataset

We now apply the method outlined in the previous sections to the MS data set. We first have to define a set of equally-spaced time points t_1, \dots, t_k . We chose 11 time points (days 100, 200, \dots , 1000 and 1095), so that each interval covers roughly 100 days (the last one covers only 95 days, since the trial ended on day 1095). This is close to the scheduled visits of the patients at the clinic, so that one unit of time corresponds roughly to the time interval between visits.

Table 5.1 shows the estimated effects of the baseline covariates on the event counts. There is a significant treatment effect for the high dose treatment

(patients in the high dose group have about $\exp(-0.40) \approx 67\%$ the number of exacerbations of a placebo patient with the same covariates at the same time point). There is some indication of a treatment effect for the low dose treatment as well, but it is not significant at the 5% level. The mean number of exacerbations decreases over time with the number of exacerbations being reduced by about 30% for each succeeding 100-day interval. Older patients tend to have fewer exacerbations. The other effects are not significant, although there is some indication that female patients and patients with a higher EDSS have more exacerbations.

Table 5.2 compares the naive, robust and jackknifed standard errors (s.e.). The comparison of the robust to the naive s.e.'s shows how different the conclusions about the covariates would be if they were based on the naive model. The ratio of the naive s.e.'s to the robust s.e.'s shows that the naive s.e.'s for most covariates are about 70% as large as the robust ones. This ratio is quite consistent across different covariates, with the exception of the covariate time. The jackknifed s.e.'s are almost identical to the robust ones, which confirms the results for the robust s.e.'s.

Table 5.3 shows the estimated effects for the mean exacerbation lengths; no covariate has a significant effect on the lengths of exacerbations, and the z-scores are all quite small. This is a bit surprising, since we have seen that the Markov chain and alternating renewal process models considered earlier showed at least some indication of covariate effects. Table 5.4 shows the estimated standard errors for the mean exacerbation lengths. Except for the covariate time, the ratio of naive to robust s.e. is roughly 25 %, which means that use of the naive s.e.'s could be quite misleading. We also see that the jackknife s.e.'s are not consistent with either the naive or the robust s.e.'s; they are even smaller than the naive ones and the ratio of the naive to the jackknifed s.e.'s varies considerably across different covariates. We anticipated that the jackknife might not work in this situation, and the results indicate that this is in fact the case.

5.8 Discussion

The analysis for the counts of exacerbations showed the anticipated results: the robust standard errors are larger than the naive ones that assume in-

covariate	estimate	robust s.e.	robust z-score
low dose	-0.0192	0.4451	-0.04
high dose	-0.0100	0.5271	-0.02
age/30	0.1936	0.9447	0.20
sex	0.0107	0.4640	0.02
time since onset	-0.0090	0.0386	-0.23
EDSS	0.0290	0.1530	0.19
time	0.0671	0.0711	0.94

Table 5.3: Estimated effects for the mean exacerbation lengths

covariate	naive s.e.	robust s.e.	jackknifed s.e.	naive/robust
low dose	0.1129	0.4451	0.0916	0.25
high dose	0.1209	0.5271	0.0867	0.23
age/30	0.2254	0.9447	0.1881	0.24
sex	0.1091	0.4640	0.0856	0.24
time since onset	0.0095	0.0386	0.0065	0.25
EDSS	0.0399	0.1530	0.0322	0.26
time	0.0645	0.0711	0.0400	0.91

Table 5.4: Estimated standard errors for the mean exacerbation lengths

dependence of data within individuals and the jackknife standard errors are comparable to the robust ones. For the length of exacerbations the naive standard errors are much smaller than the robust ones (roughly by a factor of 4 for most covariates). This means that inference based on the naive s.e.'s would be quite misleading.

We have in our approach by-passed a model for the covariances between observations on an individual. A different option would be to work with correlations induced by random effects instead of the empirical estimates (5.2). We did not pursue this approach here; see Nadeau and Lawless [22] and Ng and Cook [24].

Chapter 6

Discussion

6.1 Conclusions for the Data Set

We first compare the conclusions we have obtained by using the different models for the data set. These models address different outcomes: our Markov chain model expresses covariate effects in terms of the transition probabilities at a set of discrete time-points; alternating renewal processes express the effects in terms of transition probabilities that are continuous in time, and marked point processes model mean exacerbation counts and mean exacerbation lengths. However, these outcomes are qualitatively related in that we can identify factors that lead to qualitatively different disease processes. A higher transition probability into exacerbation is not desirable for the patient, nor is a higher count of exacerbations. A higher probability to stay in exacerbation is undesirable as well, as is a higher mean exacerbation length. Hence the z-scores from these different models can be related, as they identify covariates that lead to significantly “worse” or “better” disease processes.

For the alternating renewal process model and the marked point process model, we have obtained z-scores that are valid if we assume that data within individuals is independent (the z-scores from the fixed-effects model for the alternating renewal process model and those corresponding to the naive standard errors for the marked point process model). Although these models are clearly too simple, they were the basis for the models that we then developed to take the correlation of observations on an individual into account. Hence it is interesting to see how the z-scores from these simple models compare before

one compares the z-scores from the models that take the longitudinal nature of the data into account. Although the Markov chain fixed effects model differs in that it does not assume the independence of data within an individual, we include those results in this naive comparison. Table 6.1 shows the z-scores for the fixed effects model for the Markov chain and alternating renewal process model and the z-scores based on the naive standard errors for the marked point process model (so these are the z-scores for a marked Poisson process). Note that the renewal process model has a different interpretation for the sign of the z-score for state 2: here we model the transition from exacerbation to no exacerbation, so that a positive z-score indicates a beneficial effect for the patient. The Markov chain model models the probability of remaining in exacerbation and in the Poisson process model we model the length of exacerbations; in both cases a negative z-score indicates a beneficial effect for the patient. For the renewal process model we list the z-scores for the linear parameterization of the time-scales since we did not consider more complicated parameterizations for time in the other models. Table 6.1 shows that the z-scores for a covariate are fairly comparable (in the sense outlined earlier) across these models for state 1 as well as for state 2. An exception for state 1 is gender, which is significant in the Markov chain model but not in the other models. For state 2, the only significant (or almost significant) z-scores are for age and time since onset in the alternating renewal process model.

Table 6.2 shows the corresponding z-scores for the random effects models (Markov chains, alternating renewal processes) and the marked point processes (with robust standard errors). For state 1, the z-scores for different covariates are fairly comparable across the different models. For state 2, we see that the z-scores for the point process model are much smaller in absolute value than the corresponding z-scores from the Markov chain and renewal process models. The z-scores from the Markov chain model and the renewal process model indicate comparable estimated effects of a covariates on the disease process in both models.

The magnitude of the z-scores for the covariate effects for state 1 drops considerably from Table 6.1 to Table 6.2, but most of the effects that were significant in the simpler models are still significant in the more complicated models. For state 2, on the other hand, the effects are quite weak in the simple

covariate	Markov chain (fixed effects)	renewal process (fixed effects)	point process (naive s.e.)
for state 1			
low dose	-2.20	-2.53	-2.09
high dose	-5.25	-4.77	-4.98
age	-3.43	-2.84	-2.84
gender	2.62	1.37	1.84
time since onset	-0.55	-1.05	-0.70
EDSS	3.32	2.65	1.97
for state 2			
low dose	-0.90	0.29	-0.17
high dose	-0.40	0.17	-0.08
age	1.01	-1.90	0.86
gender	-0.31	-0.15	0.10
time since onset	-1.30	1.95	-0.94
EDSS	0.97	-1.06	0.73

Table 6.1: Z-scores for different simple models (fixed-effects and naive standard errors)

covariate	Markov chains (random intercept)	renewal process (random intercept)	point process (robust s.e.)
for state 1			
low dose	−1.93	−1.95	−1.45
high dose	−3.57	−3.57	−3.60
age	−1.93	−2.25	−2.03
gender	1.72	1.14	1.24
time since onset	−0.55	−0.78	−0.70
EDSS	1.86	1.74	1.57
for state 2			
low dose	−0.89	0.13	−0.04
high dose	−0.24	−0.25	−0.02
age	0.80	−0.99	0.20
gender	−0.09	−0.45	0.02
time since onset	−0.74	0.76	−0.23
EDSS	0.92	−1.21	0.19

Table 6.2: Z-scores for different models (random effects and robust standard errors)

models and there is no significant covariate in any of the more complicated models. Both tables show reasonable overall agreement in the assessment of covariates.

6.2 Comparison of Methods

In the introduction, we have already compared the modeling aspects of the three different models. We also discussed the interpretation of covariates in these models. Here we consider the practical aspects and usefulness of these models.

“All models are wrong but some are useful.”

Following this famous quote by G.E.P. Box, we want to compare these different models from the perspective of an applied statistician:

- How much of the structure can the model capture?
- How helpful is the model for the assessment of covariates?
- How difficult is the estimation of parameters?

We have already discussed the assessment of covariates at some length in Chapter 1. Here we focus primarily on the estimation aspects and make some general remarks on each model.

We have seen that Markov chains model the exacerbation data by modeling the transition probabilities. This means that the effects of covariates are judged conditionally on previous responses and the exacerbation data enters into the analysis in the form of transition counts. As discussed in Chapter 1, the resulting interpretation of covariates is not straightforward. We have also seen that this model cannot deal effectively with long-range dependence of data over time; this forced us to consider a subset of the available daily data. Even then, it turned out that a first-order model is not quite appropriate, and the interpretation of covariates in higher-order models is even less clear.

The estimation of parameters in Markov chain models is straightforward. We can use generalized linear models software for fixed effects models

and we have the MIXOR software available for random effects models, although MIXOR could not handle complicated models.

Alternating renewal processes incorporate the length of exacerbations directly into the model. This takes full advantage of the information available and makes the interpretation of covariates similar to that in survival analysis. For our application, however, it turned out that the estimated effects were comparable (judged by significance) to those obtained in the Markov chain model, so that the extra effort did not pay off in this sense. But this might be due to highly variable behaviour of the times in exacerbation, so that the length of exacerbations is not related to any of the covariates.

The estimation for the fixed effects model with simple parameterizations for time as a covariate was not difficult; the random effects models turned out to be very difficult to fit. The estimation procedure suggested by Ng and Cook [23] was numerically unstable, and even with our modification of the essential integration the estimation was very slow. These problems might be overcome with the use of more sophisticated software, but this may not be worth the effort in our application.

Our marked point processes approach models the event counts and length of exacerbations for a fixed set of time-points; Lawless and Nadeau [18] indicate that this can be extended to a continuous time setting. The GEE approach allows us to use estimates obtained under the Poisson assumptions in more general contexts and makes the interpretation of the coefficients straightforward. The results for the counts of exacerbations were very consistent with the results from the other models, and the jackknife s.e.'s were almost identical to the robust ones. For the exacerbation length, on the other hand, the robust z-scores were quite small in comparison to the other models.

Computations for the marked point processes were straightforward; the log-linear modeling of the mean functions enabled us to use GLM-software and also simplified the calculations for the sandwich estimator.

6.3 Conclusion

Our main objective was to incorporate the length of exacerbations into the analyses (along with the length of exacerbation-free time periods) to assess the effect of covariates on this two-state process. Different models approach this objective quite differently, but in all models the estimated effects on the length of exacerbations are quite weak. We have also seen that the second state seems to behave quite differently from the first. One of the problems all these models face is that they have to model the length of an exacerbation conditional on an exacerbation occurring. This means that a person without any exacerbation does not contribute to the inference for the length of exacerbations. Together with the fact that the number of exacerbations decreases over time, this means that it will be difficult to estimate a treatment effect. For example, if the treatment works well, the patient will have fewer exacerbations and hence contribute less to the inference for the lengths of exacerbations. But without plausible models for the relationship between number and lengths of exacerbations, this would be difficult to incorporate into any analysis.

We mentioned earlier that we also have the severity of exacerbations available. Some of the analyses we suggested could also incorporate this information. Markov chains can (at least in theory) be extended to any number of states, not just the two we utilized. The alternating renewal process could also be extended to include more states. Marks in marked point processes can also be categorical; if we just used the severity as marks we would in fact simply obtain a multivariate point process. While all of these extensions or alternatives are possible in principle, the estimation could be more difficult, particularly for the alternating renewal process model.

This thesis suggests that the covariates available in this data set do not contain much information about the lengths of exacerbations. This could mean either that there really is little information in the lengths of exacerbations and they are fairly random, or that we need more specific covariates that are not as crude as gender, for example. In any case, a better understanding of the biological mechanisms underlying the exacerbations would be helpful to answer these questions conclusively.

Bibliography

- [1] Aalen, O.O. and E. Husebye (1991). Statistical analysis of repeated events forming renewal processes. *Statistics in Medicine* **10**, 1227-1240.
- [2] Anderson, T.W. and L. Goodman (1957). Statistical inference about markov chains. *Annals of Mathematical Statistics* **28**, 89-110.
- [3] Cook, R.J. and E.T.M. Ng (1997). A logistic-bivariate normal model for overdispersed two-state Markov processes. *Biometrics* **53**, 358-364.
- [4] Cox, D.R. (1972). The statistical analysis of dependencies in point processes. In *Stochastic point processes*, Lewis, P.A.W. (ed.). John Wiley, New York, pp. 55-66.
- [5] Cox, D.R. and V. Isham (1980). *Point processes*. Chapman and Hall, London.
- [6] D'yachkova, Y. (1997). *Analysis of longitudinal data from the betaseron multiple sclerosis clinical trial*. M.Sc. Thesis, Department of Statistics, University of British Columbia.
- [7] Diggle, P.J., K.Y. Liang and S.L. Zeger (1994). *Analysis of longitudinal data*. Clarendon Press, Oxford.
- [8] European Study Group on Interferon Beta-1b in Secondary Progressive MS (1998). Placebo-controlled multicentre randomised trial of interferon beta-1b in treatment of secondary progressive multiple sclerosis. *The Lancet* **352**, 1491-1497.
- [9] The IFNB Multiple Sclerosis Study Group (1993). Interferon beta-1b is effective in relapsing remitting multiple sclerosis: clinical results of a multi-

- center, randomized, double-blind, placebo-controlled trial. *Neurology* **44**, 665-661.
- [10] The IFNB Multiple Sclerosis Study Group (1995). Interferon beta-1b in the treatment of multiple sclerosis: final outcome of the randomized controlled trial. *Neurology* **45**, 1277-1285.
 - [11] Jacobs, J.D. et al. (1996). Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. *Annals of Neurology* **39**, 285-294.
 - [12] Johnson, K.P. et al. (1995). Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis. *Neurology* **45**, 1268-1276.
 - [13] Kalbfleisch, J.D. and J.F. Lawless (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association* **80**, 863-871.
 - [14] Kingman, J.F.C. (1993). *Poisson processes*. Clarendon Press, Oxford.
 - [15] Lawless, J.F. (1982). *Statistical models and methods for lifetime data*. John Wiley and Sons.
 - [16] Lawless, J.F. (1987). Regression methods for Poisson process data. *Journal of the American Statistical Association* **82**, 808-815.
 - [17] Lawless, J. F. (1995). The analysis of recurrent events for multiple subjects. *Applied Statistics* **44**, 487-498.
 - [18] Lawless, J.F. and C. Nadeau (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37**, 158-168.
 - [19] Lawless, J.F. and K. Thiagarajah (1996). A point-process model incorporating renewals and time trends, with application to repairable systems. *Technometrics* **38**, 131-138.
 - [20] Lindeboom, M. and G.J. Van Den Berg (1994). Heterogeneity in models for bivariate survival: the importance of the mixing distribution. *Journal of the Royal Statistical Society, Series B* **56**, 49-60.

- [21] Muenz, L.R. and L.V. Rubinstein (1985). Markov models for covariate dependence of binary sequences. *Biometrics* **41**, 91-101.
- [22] Nadeau, C. and J.F. Lawless (1998). Inference for means and covariances of point processes through estimating functions. *Biometrika* **85**, 893-906.
- [23] Ng, E.T.M. and R.J. Cook (1997). Modeling two-state disease processes with random effects. *Lifetime Data Analysis* **3**, 315-335.
- [24] Ng, E.T.M. and R.J. Cook (1999). Robust inference for bivariate point processes. *Canadian Journal of Statistics* **27**, 509-524.
- [25] Oakes, D. and L. Cui (1993). On semiparametric inference for modulated renewal processes. *Biometrika* **81**, 83-90.
- [26] Press, W.H. et al. (1988). *Numerical Recipes in C*, Cambridge University Press.
- [27] PRISMS Study Group (1998). Randomised double-blind placebo-controlled study of interferon beta-1a in relapsing-remitting multiple sclerosis. *The Lancet* **352**, 1498-1504.
- [28] Saerndal, C.E., B. Swensson and J. Wretman (1992). *Model assisted survey sampling*, Springer-Verlag, New York.
- [29] Smith, A. (1999). *Design strategies for repeated MRI scanning in multiple sclerosis clinical trials*. M.Sc. Thesis, Department of Statistics, University of British Columbia.
- [30] Stiratelli, R., N. Laird and J.H. Ware (1984). Random effects models for serial observations with binary response. *Biometrics* **40**, 961-972
- [31] Stoer, J. and R. Bulirsch (1980). *Introduction to numerical analysis*, Springer-Verlag, New York.
- [32] Venables, W. N. and B.D. Ripley (1997). *Modern applied statistics with S-PLUS, 2nd ed.*, Springer-Verlag, New York.
- [33] Venables, W. N. and B.D. Ripley (1997). *Modern applied statistics with S-PLUS, 2nd ed.*, online complements. Springer-Verlag, New York.

- [34] Xue, X. and R. Brookmeyer (1996). Bivaraiate frailty model for the analysis of multivariate survival time. *Lifetime Data Analysis* **2**, 277-289.